

# **Epigenetics of Crohn's Disease, Ulcerative Colitis, and Phenotypically Normal Individuals**

By

© 2017

Delisa L Phillips

M.S., University of New Mexico, 2009

M.A., University of Arkansas, 2005

B.A., University of Arkansas, 2003

Submitted to the graduate degree program in Anthropology and the Graduate Faculty of the  
University of Kansas in partial fulfillment of the requirements for the degree of Doctor of  
Philosophy.

---

Chair: Michael Crawford

---

Justin Blumenstiel

---

Kristin Young

---

Jennifer Raff

---

Bartholomew Dean

Date Defended: 2 May 2017

The dissertation committee for Delisa L Phillips certifies that this is the approved version of the following dissertation:

**Epigenetics of Crohn's Disease, Ulcerative Colitis, and  
Phenotypically Normal Individuals**

---

Chair: Michael Crawford

Date Approved: 2 May 2017

## ***ABSTRACT***

### *Background*

It is difficult to distinguish between Crohn's disease and ulcerative colitis, as they share many symptoms and many susceptibility genetic loci. Due to the similarities in phenotype between Crohn's disease and ulcerative colitis, many patients are misdiagnosed for years. For this project, I collected genetic material, medical histories, and environmental variables from individuals with Crohn's disease, individuals with ulcerative colitis, and individuals that are phenotypically normal. The purpose of this project is to determine the differential epigenetic methylation of genes in individuals with Crohn's disease and ulcerative colitis compared to phenotypically normal controls.

### *Materials and Methods*

Participants were recruited for this project at the Kansas Medical Center. Twenty of these individuals had Crohn's disease, ten had ulcerative colitis, and thirty-one were phenotypically normal. Buccal swabs and information relevant to this project (via a questionnaire) was collected from the individuals. DNA was extracted then bisulfite-converted so that the unmethylated cytosines would be converted to uracils. Real-time PCR was performed on the promoter regions of three genes both prior to and post-bisulfite conversion, as well as on the regions of the genes containing known associated SNPs.

The three genes of interest in this project are the *NOD2*, *ATG16L1*, and *PTPN2* genes. Each of these genes has been found to be associated with Crohn's disease. These three genes work in concert by identifying pathogenic bacteria, forming a response, and ultimately clearing out the bacteria.

## *Results*

The statistical tests used in this project showed a few significant differences between Crohn's disease, ulcerative colitis, and phenotypically normal samples from the questionnaire. From the bisulfite-converted samples, however, there were no obvious differences between the three groups.

## *ACKNOWLEDGEMENTS*

I would like to thank the patients from the Kansas City Medical Center who volunteered for this study, as well as the wonderful doctors who allowed me to collect samples from their patients: Dr. John Bonino, Dr. Rajeswari Anaparthi, and Dr. Daniel Buckles. I would like to thank the members of my committee who have all worked so hard to insure that I succeed: Dr. Michael Crawford, Dr. Kristin Young, Dr. Justin Blumenstiel, Dr. Jennifer Raff, and Dr. Bartholomew Dean. I would also like to thank my husband and son for putting up with me through all of my moodiness and stress during the writing of this dissertation.

I would like to dedicate this to my parents, Dale and Dee Phillips. The two of you have been my rock throughout my entire academic career. Over the years, you have provided me with a place to live, food to eat, financial support as needed, and most importantly, the emotional support that I needed to complete my Ph.D. You have made me the woman I am today, and I am thankful for the two of you every day of my life.

This project was partially funded by the Carroll Clark Award.

## Table of Contents

<i>ABSTRACT</i> .....	<i>III</i>
<i>ACKNOWLEDGEMENTS</i> .....	<i>V</i>
<i>TABLE OF FIGURES</i> .....	<i>VII</i>
<b>CHAPTER 1: INTRODUCTION</b> .....	<b>1</b>
<b>CHAPTER 2: LITERATURE REVIEW</b> .....	<b>3</b>
<i>HISTORY OF CROHN’S DISEASE</i> .....	<i>3</i>
<i>AUTOIMMUNE DISEASES</i> .....	<i>7</i>
<i>WHY STUDY AUTOIMMUNE DISEASES?</i> .....	<i>9</i>
<i>GENETIC CONTRIBUTION TO AUTOIMMUNE DISEASES</i> .....	<i>11</i>
<i>NORMAL FUNCTIONING OF THE IMMUNE SYSTEM</i> .....	<i>20</i>
<i>BACTERIAL/VIRAL THEORY OF CROHN’S DISEASE</i> .....	<i>41</i>
<i>GENETICS OF CROHN’S DISEASE</i> .....	<i>48</i>
<i>HUMAN EPIGENETICS</i> .....	<i>82</i>
<b>CHAPTER 3: MATERIALS &amp; METHODS</b> .....	<b>128</b>
<i>SUBJECT SELECTION CRITERIA AND SAMPLE SIZE JUSTIFICATION</i> .....	<i>130</i>
<i>DNA EXTRACTION FROM BUCCAL SWABS</i> .....	<i>132</i>
<i>PRIMER DESIGN</i> .....	<i>133</i>
<i>METHODS FOR DETECTING DNA METHYLATION</i> .....	<i>135</i>
<i>POLYMERASE CHAIN REACTION (PCR)</i> .....	<i>139</i>
<i>DATABASE</i> .....	<i>141</i>
<i>STATISTICAL TESTS</i> .....	<i>142</i>
<b>CHAPTER 4: RESULTS</b> .....	<b>145</b>
<i>PCR RESULTS</i> .....	<i>145</i>
<i>STATISTICAL RESULTS</i> .....	<i>146</i>
<i>MEANS AND STANDARD DEVIATIONS</i> .....	<i>146</i>
<i>T-TESTS</i> .....	<i>148</i>
<i>CORRELATIONS AND FISHER’S EXACT TESTS</i> .....	<i>150</i>
<b>CHAPTER 5: DISCUSSION AND CONCLUSIONS</b> .....	<b>154</b>
<b>REFERENCES</b> .....	<b>158</b>
<b>APPENDICES</b> .....	<b>175</b>
<i>APPENDIX 1 – IRB CONSENT FORM</i> .....	<i>175</i>
<i>APPENDIX 2 – DATA COLLECTION WORKSHEET</i> .....	<i>183</i>
<i>APPENDIX 3 – COUNT OF EACH VARIABLE FROM QUESTIONNAIRE</i> .....	<i>186</i>
<i>APPENDIX 4 – AMPLIFICATION PLOTS AND MELT CURVES</i> .....	<i>188</i>
<i>APPENDIX 5 – INTERVAL PLOTS POST-PCR</i> .....	<i>197</i>
<i>APPENDIX 6: CONTINGENCY TABLES WITH ODDS RATIOS, RELATIVE RISKS, AND CONFIDENCE INTERVALS</i> .....	<i>200</i>

## TABLE OF FIGURES

FIGURE 1 INCIDENCE OF IMMUNE DISORDERS FOR FOUR AUTOIMMUNE DISEASES. (BACH). THE X-AXIS DENOTES THE YEARS 1950-2000. THE Y-AXIS DENOTES THE PERCENT INCREASE OF IMMUNE DISORDERS. CROHN'S DISEASE HAS THE LARGEST INCREASE. THE DATA FOR THIS GRAPH WAS OBTAINED FROM FOUR SEPARATE ARTICLES (DUBOIS ET AL. 1998; PUGLIATTI ET AL. 2001; SWARBRICK ET AL. 2001; TUOMILEHTO ET AL. 1999).	10
TABLE 1 AUTOIMMUNE CONCORDANCE RATES IN TWINS	14
FIGURE 2 FEMALE TO MALE RATIO OF SELECTED AUTOIMMUNE DISEASE EXPRESSED AS PERCENTAGE (X AXIS) AND CALCULATED AS AVERAGE (INVERNIZZI ET AL. 2009). THE POPULATIONS STUDIED WERE IN THE UNITED STATES AND DENMARK. INFLAMMATORY BOWEL DISEASES INCLUDE CROHN'S DISEASE AND ULCERATIVE COLITIS.	17
TABLE 2 MEAN AGE OF DIAGNOSIS FOR SIX AUTOIMMUNE DISEASES.	19
FIGURE 3 THE COMPLEMENT PATHWAY (DUNKELBERGER AND SONG).	22
TABLE 3 ENVIRONMENTAL AND NUTRITIONAL FACTORS THAT AFFECT CROHN'S DISEASE.	49
FIGURE 4 GENE ONTOLOGY FUNCTIONAL ENRICHMENT ANALYSIS OF DIFFERENTIALLY EXPRESSED GENES IN CROHN'S DISEASE. THE X-AXIS IS THE NUMBER OF GENES THAT ARE DIFFERENTIALLY EXPRESSED. THE Y-AXIS SHOWS THE INCREASE IN P VALUE FROM BOTTOM TO TOP (LIN ET AL. 2014).	52
FIGURE 5 NOD2 SIGNALING PATHWAYS (KASER AND BLUMBERG).	55
TABLE 4 THE CONTINENT AND POPULATIONS STUDIED IN THE RESEARCH BY NAKAGOME ET AL. (NAKAGOME ET AL.), AS WELL AS THE SAMPLE SIZE OF EACH POPULATION USED IN THE STUDY.	58
FIGURE 6 GENE TREES FOR H1, H2, AND H3 HAPLOGROUPS IN EUROPEANS. THE HEIGHTS OF THE TREES ARE PROPORTIONAL TO THE TIME TO THE MOST COMMON RECENT ANCESTOR. BLACK CIRCLES INDICATE MUTATIONS SHARED AMONG AFRICANS AND EUROPEANS, WHEREAS STARS INDICATE EUROPEAN-SPECIFIC MUTATIONS. NUMBERS ON THE BLACK CIRCLES OR STARS CORRESPOND TO THE SNP# (NAKAGOME ET AL. 2012).	60
FIGURE 7 HAPLOTYPE NETWORK IN EUROPEANS BASED ON 38 SNP SITES. THE SIZE OF THE CIRCLE CORRESPONDS TO THE FREQUENCY OF THE HAPLOTYPE. THE NUMBERS ON THE BRANCHES CORRESPOND TO THE SNP#. THE STARS ON THE BRANCHES INDICATE EUROPEAN-SPECIFIC ALLELES. UNDERLINED SNP SITE NUMBERS INDICATE THE SITES THAT THE PHASE INFERENCE FAILED TO ASSIGN TO ONE OF TWO CHROMOSOMES IN AN INDIVIDUAL (NAKAGOME ET AL. 2012).	60
TABLE 5 THE SNP, LOCATION, RISK ALLELE, AND ODDS RATIO ASSOCIATED WITH THE SNPS PREVIOUSLY FOUND TO BE ASSOCIATED WITH THE THREE GENES USED IN THIS PROJECT (BARRETT ET AL. 2008).	75
TABLE 6 LIST OF GENES WITH EITHER SNP OR METHYLATION PROFILE ASSOCIATED WITH CROHN'S DISEASE.	80
FIGURE 8 "ROLES FOR EPIGENETICS IN IBD PATHOGENESIS. EPIGENETICS MAY INTERACT WITH GENETIC FACTORS (1), ENVIRONMENTAL FACTORS LIKE NUTRITION, SMOKING, STRESS OR HYGIENE, AND THE INTESTINAL MICROBIOME (3) IN AFFECTING THE IMMUNE SYSTEM. PHAGOCYtic CELLS WITHIN THE LAMINA PROPRIA (E.G., MONOCYTES/MACROPHAGES, DCs, NEUTROPHILS) AND EPITHELIAL CELL BARRIER REPRESENT THE CENTRAL COMPONENTS OF THE INTESTINAL INNATE IMMUNE SYSTEM (4). ANTIGENS REACHING THE LAMINA PROPRIA ACTIVATE INNATE IMMUNE CELLS FOLLOWED BY A RESPONSE OF T LYMPHOCYTES (Th, T-HELPER CELLS; TREG, T-REGULATORY CELLS), WHICH REPRESENT THE KEY CELL POPULATION OF THE ADAPTIVE IMMUNE SYSTEM (5). THE INNATE AND ADAPTIVE IMMUNITY ARMS ARE TIGHTLY CROSS-REGULATED SERVING TO UPHOLD INTESTINAL HOMEOSTASIS AND THUS TO CONTROL THE COMPLEX COMMENSAL-HOST CROSSTALK (6). THE SUBSEQUENT IMMUNE RESPONSE HAS CONSEQUENCES ON THE INITIATION, RESOLUTION, AND/OR PROGRESSION OF INTESTINAL INFLAMMATION (7). BACTERIA CAN ALSO REGULATE THE EPITHELIAL GENE EXPRESSION THROUGH EPIGENETIC MECHANISMS (8)" (DABRITZ AND MENHENIOTT 2014).	85
TABLE 7 LIST OF HISTONE MODIFICATIONS AND THEIR EFFECTS ON GENE EXPRESSION, TRANSCRIPTION, AND DNA REPAIR.	98
FIGURE 9 A SELF-REINFORCING LOOP LINKING siRNAs TO DNA AND HISTONE METHYLATION IN ARABIDOPSIS THALIANA "ELABORATE FEEDBACK BETWEEN SMALL RNAs AND DNA AND HISTONE METHYLATION UNDERLIES A ROBUST SILENCING PATHWAY AT SITES OF ASYMMETRICAL CYTOSINE METHYLATION IN THE ARABIDOPSIS THALIANA GENOME. TWO PLANT-SPECIFIC POLYMERASES TRANSCRIBE THE CRITICAL RNAs. RNA POLYMERASE IV (POL IV) TRANSCRIPTS ARE PROCESSED BY THE RNA-DEPENDENT RNA POLYMERASE RDR2 AND THE DICER PROTEIN DCL3 INTO 24-NUCLEOTIDE (NT) SMALL INTERFERING (siRNAs), WHILE POL V TRANSCRIPTS ACT AS THEIR TARGETS. THE ARGONAUTE PROTEIN AGO4, THE siRNA-DEPENDENT RECRUITMENT OF WHICH TO POL V TRANSCRIPTS IS REINFORCED BY INTERACTIONS WITH THE GW DOMAINS OF POL V AND AN ASSOCIATED ELONGATION FACTOR KTF1, IN TURN RECRUITS THE CHH DNA METHYLTRANSFERASE DRM2. RDM1 ASSOCIATES WITH THE POL V-AGO4-DRM2 COMPLEX AND MAY LINK siRNA AMPLIFICATION TO PRE-EXISTING DNA METHYLATION. MEANWHILE, ANOTHER DNA METHYLTRANSFERASE THAT TARGETS CHG SITES	

FOR MAINTENANCE, CMT3, IS RECRUITED DIRECTLY TO METHYLATED HISTONE H3 LYSINE 9 (H3K9). SILENCING BY DNA METHYLATION IS AUGMENTED BY H3K9 METHYLATION, WHICH IS DEPOSITED BY THE ENZYMES KYP, SUVH5, AND SUVH6. THESE METHYLATION EVENTS ARE COUPLED TO ONE ANOTHER AND TO siRNA ACTIVITY IN SEVERAL WAYS. KYP IS RECRUITED DIRECTLY TO METHYLATED DNA, WHERE IT METHYLATES NEIGHBORING HISTONES, AND THE H3K9 METHYLATION READER SHH1 RECRUITS POL IV TO PROMOTE siRNA GENERATION, WHILE THE DNA METHYLATION READERS SUVH2 AND SUVH9 RECRUIT POL V TO PROMOTE AGO4 TARGETING AND FURTHER DNA METHYLATION. THUS, THE DIFFERENT METHYLATION READERS, RNA POLYMERASES AND AGO4 ACT TOGETHER TO CREATE SELF-REINFORCING INTERACTIONS BETWEEN PRE-EXISTING DNA METHYLATION AND siRNA AMPLIFICATION. ERASURE OF DNA METHYLATION BY MUTATIONS IN EITHER THE HISTONE DEACETYLASE HDA6 OR THE MAINTENANCE DNA METHYLTRANSFERASE MET1 RESULTS IN LOSS OF siRNA BIOGENESIS, EMPHASIZING THE IMPORTANCE OF THESE SELF-ENFORCING INTERACTIONS” (HOLOCH AND MOAZED 2015).	101
FIGURE 10 CLUSTER ANALYSIS OF 14 CROHN’S DISEASE-ASSOCIATED METHYLATION CpG SITES IN B CELL LINES. LEFT PANEL HIERARCHICAL CLUSTERING HEAT MAP OF METHYLATION DATA. RIGHT PANEL DATA ANALYSIS OF THE 14 CD-ASSOCIATED LOCI FROM B CELLS OF IBD (CD AND UC) PATIENTS AND NON-IBD SIBLINGS. GENE SYMBOLS CONTAINED WITHIN THE TARGET ID BEFORE THE FIRST UNDERScore. ROWS CORRESPOND TO CpG SITES. COLOR INDICATION FOR SAMPLES: ORANGE REPRESENTS B CELLS FROM IBD PATIENTS, AND LIGHT BLUE REPRESENTS NON-IBD SIBLINGS; ROWS CORRESPOND TO CpG SITES; B-VALUE: RED INDICATES INCREASED METHYLATION, AND BLUE INDICATES DECREASED METHYLATION. RIGHT PANEL: D_MEAN = MEAN B-VALUE (FRACTIONAL METHYLATION FROM 0 TO 1) OF DISEASED TISSUE; D_SD = STANDARD DEVIATION OF DISEASED TISSUE; N_MEAN = MEAN B-VALUE OF NORMAL TISSUE; N_SD = STANDARD DEVIATION OF DISEASED TISSUE; DIFFER = ABSOLUTE MEAN DIFFERENCE BETWEEN DISEASED AND NORMAL TISSUES. P = P VALUE (LIN ET AL. 2012). THE HEAT MAP ON THE LEFT SHOWS THAT THERE IS INCREASED METHYLATION IN INDIVIDUALS WITH IBD FOR MOST OF THE GENES TESTED, WHEREAS THEIR HEALTHY SIBLINGS PRIMARILY SHOWED A DECREASE IN METHYLATION.	121
TABLE 8 A SUMMARY OF THE CURRENT RESEARCH ON DIFFERENTIAL METHYLATION IN INDIVIDUALS WITH CROHN’S DISEASE, AS COMPARED TO CONTROLS. TABLE INCLUDES THE AUTHORS OF THE PAPER, THE NUMBER OF CROHN’S DISEASE PATIENTS, ULCERATIVE COLITIS, AND CONTROLS USED IN THE STUDY. ALSO INCLUDED IS THE TYPE OF SAMPLE COLLECTED, THE GENES WITH DIFFERENTIALLY METHYLATED LOCI, AND THE NUMBER OF LOCI FOUND TO HAVE METHYLATION.	124
TABLE 9 THE PRIMER SEQUENCES FOR THE THREE GENES FOLLOWING BISULFITE CONVERSION. THE CpG SITES ARE IN BOLD, AND THE CONVERTED CYTOSINES ARE LISTED AS A CAPITAL T IN THE FORWARD PRIMER AND A CAPITAL A IN THE REVERSE PRIMER.	134
TABLE 10 NUMBER OF SAMPLES FOR EACH MEASUREMENT BROKEN DOWN BY CROHN’S DISEASE, ULCERATIVE COLITIS, AND PHENOTYPICALLY NORMAL.	145
TABLE 11 MEAN, STANDARD DEVIATION, AND NUMBER OF SAMPLES FOR NOD2 RESULTS. THE RESULTS ARE SEPARATED INTO THOSE INDIVIDUALS WITH CROHN’S DISEASE, THOSE WITH ULCERATIVE COLITIS, AND PHENOTYPICALLY NORMAL CONTROLS.	146
TABLE 12 MEAN, STANDARD DEVIATION, AND NUMBER OF SAMPLES FOR ATG16L1 RESULTS. THE RESULTS ARE SEPARATED INTO THOSE INDIVIDUALS WITH CROHN’S DISEASE, THOSE WITH ULCERATIVE COLITIS, AND PHENOTYPICALLY NORMAL CONTROLS.	146
TABLE 13 MEAN, STANDARD DEVIATION, AND NUMBER OF SAMPLES FOR PTPN2 RESULTS. THE RESULTS ARE SEPARATED INTO THOSE INDIVIDUALS WITH CROHN’S DISEASE, THOSE WITH ULCERATIVE COLITIS, AND PHENOTYPICALLY NORMAL CONTROLS.	147
TABLE 14 T TEST COMPARING CROHN’S DISEASE SAMPLES TO NON-CROHN’S DISEASE SAMPLES FOR THE NOD2 GENE. SIGNIFICANT P-VALUES ARE IN RED.	148
TABLE 15 T TEST COMPARING ULCERATIVE COLITIS SAMPLES TO NON-UC SAMPLES FOR THE NOD2 GENE, SIGNIFICANT P-VALUES ARE IN RED.	149
TABLE 16 T TEST COMPARING IBD SAMPLES TO NON-IBD SAMPLES FOR THE NOD2 GENE. SIGNIFICANT P-VALUES ARE IN RED.	149
TABLE 17 CORRELATION BETWEEN EACH OF THE NOD2 VARIABLES AND CROHN’S DISEASE. SIGNIFICANT P-VALUES ARE IN RED.	150
TABLE 18 FISHER’S EXACT TEST COMPARING CD TO VARIABLES. SIGNIFICANT P-VALUES <0.0042 ARE IN RED.	150
TABLE 19 CORRELATIONS BETWEEN EACH OF THE NOD2 VARIABLES AND ULCERATIVE COLITIS. SIGNIFICANT P-VALUES ARE IN RED.	151
TABLE 20 FISHER’S EXACT TEST COMPARING UC TO VARIABLES. SIGNIFICANT P-VALUES <0.0042 ARE IN RED.	151



TABLE 21 CORRELATIONS BETWEEN EACH OF THE NOD2 VARIABLES AND IBD. SIGNIFICANT P-VALUES ARE IN RED.	152
TABLE 22 FISHER'S EXACT TEST COMPARING IBD TO VARIABLES. SIGNIFICANT P-VALUES <0.0042 ARE IN RED.	152
TABLE 24 COUNT OF EACH VARIABLE FROM THE QUESTIONNAIRE BROKEN DOWN BY CROHN'S DISEASE, ULCERATIVE COLITIS, AND PHENOTYPICALLY NORMAL SAMPLES.	186
FIGURE 11 AMPLIFICATION PLOT OF NOD2BC FOR CROHN'S DISEASE PATIENTS.	188
FIGURE 12 AMPLIFICATION PLOT OF NOD2BC FOR ULCERATIVE COLITIS SAMPLES.	188
FIGURE 13 AMPLIFICATION PLOT OF NOD2BC FOR PHENOTYPICALLY NORMAL SAMPLES.	189
FIGURE 14 MELT CURVE OF NOD2BC FOR CROHN'S DISEASE SAMPLES.	189
FIGURE 15 MELT CURVE OF NOD2BC FOR ULCERATIVE COLITIS SAMPLES.	190
FIGURE 16 MELT CURVE OF NOD2BC FOR PHENOTYPICALLY NORMAL SAMPLES.	191
FIGURE 17 AMPLIFICATION PLOT OF ATG16L1BC FOR CROHN'S DISEASE SAMPLES.	192
FIGURE 18 AMPLIFICATION PLOT OF ATG16L1BC FOR ULCERATIVE COLITIS SAMPLES.	192
FIGURE 19 AMPLIFICATION PLOT OF ATG16L1BC FOR PHENOTYPICALLY NORMAL SAMPLES.	193
FIGURE 20 MELT CURVE OF ATG16L1BC FOR CROHN'S DISEASE SAMPLES.	193
FIGURE 21 MELT CURVE OF ATG16L1BC FOR ULCERATIVE COLITIS SAMPLES.	194
FIGURE 22 MELT CURVE OF ATG16L1BC FOR PHENOTYPICALLY NORMAL SAMPLES.	194
FIGURE 23 AMPLIFICATION OF PTPN2BC FOR CROHN'S DISEASE SAMPLES.	195
FIGURE 24 AMPLIFICATION PLOT OF PTPN2BC FOR ULCERATIVE COLITIS SAMPLES.	195
FIGURE 25 AMPLIFICATION PLOT OF PTPN2BC FOR PHENOTYPICALLY NORMAL SAMPLES.	196
FIGURE 26 CtNOD2BC INTERVAL PLOT	197
FIGURE 27 TmNOD2BC INTERVAL PLOT	197
FIGURE 28 CtATG16L1BC INTERVAL PLOT	198
FIGURE 29 TmATG16L1BC INTERVAL PLOT	198
FIGURE 30 CtPTPN2BC INTERVAL PLOT	199
FIGURE 31 TmPTPN2BC INTERVAL PLOT	199

## CHAPTER 1: INTRODUCTION

The purpose of this project is to determine the epigenetic methylation status of genes in individuals with Crohn's disease and ulcerative colitis compared to phenotypically normal controls. The ultimate goal is to be able to determine a non-invasive method of testing for Crohn's disease. This will be accomplished by comparing methylation profiles and environmental factors. The methylation of a gene can influence the activity of that gene. A methylated gene may either upregulate or downregulate gene expression. By examining the methylation of different genes associated with Crohn's disease and ulcerative colitis, it is hoped that more insight into the onset and severity of the disease might be learned. At present, there is no definitive test for Crohn's disease. It is diagnosed primarily by ruling out other diseases. Individuals with Crohn's disease often undergo many years of tests (including blood, urine, upper GI, and small bowel series tests, as well as many colonoscopies) before they are finally diagnosed with the disease. In many cases, it is difficult to differentiate between Crohn's disease and ulcerative colitis, and the doctor will give a general diagnosis of "inflammatory bowel disease/syndrome (McDermott et al.)". Crohn's disease and ulcerative colitis exhibit many of the same symptoms, and it has been found that they share many of the same single nucleotide polymorphisms (SNPs) that are indicative of being at risk for developing the disease. This makes it difficult for individuals with Crohn's disease, as many of the medications and treatments [Mesalamine (Pentasa), azathioprine (Imuran), and infliximab (Remicade), for example] for IBS and ulcerative colitis do not work well on Crohn's disease. It is hoped that this project may be able to lead to a better, and less invasive, diagnosis tool than what is currently available.

For the purposes of this project, the promoter regions of three genes of the immune system will be analyzed. The promoter region has been found to be more likely to be differentially methylated than other regions of the genome (McDermott et al. 2016). The genes selected for intensive analysis are involved in pattern recognition of the innate immune system and apoptosis (*NOD2*), the intracellular autophagy complex (*ATG16L1*), and regulating the immune system by inhibiting T-cell proliferation and limiting defects in the epithelial barrier (*PTPN2*). The *NOD2* and *PTPN2* genes are also signaling genes within the immune system. Defects in these three genes have been shown to lead to autoimmune disorders, such as Crohn's disease (Abbas et al. 2014; Barrett and Chandra 2011; Cho and Brant 2011; Drouet et al. 2011; Fabio et al. 2011a; Festen et al. 2011; Invernizzi and Gershwin 2009; Marcil et al. 2011; Newman et al. 2009; Nimmo et al. 2011; Rodriguez-Bores 2007; Scharl et al. 2011a). This project will compare the methylation present in the promoter regions of these three genes, as well as compare various environmental factors via a questionnaire. In addition to using this questionnaire to determine risk factors, the questionnaire will also be used to account for ancestry, age, and environmental factors. Statistical tests will determine whether or not a single variable or a combination of variables may be used for developing a test for Crohn's disease. It is important to be able to control for ancestry and environmental factors, as significant differences found may be due to different backgrounds, rather than due to Crohn's disease. Based on the literature, there are significant differences in methylation patterns between Crohn's disease patients and non-symptomatic controls. It is hoped that these differences may be utilized to eventually develop a better diagnostic tool for Crohn's disease.

## CHAPTER 2: LITERATURE REVIEW

### *HISTORY OF CROHN'S DISEASE*

The first recorded instances of Crohn's disease appeared in Europe and North America during the middle part of the 20<sup>th</sup> century (Hermon-Taylor 2009). The first to describe Crohn's disease were Dr. Burrill Crohn, Dr. Leon Ginzberg, and Dr. Gordon D. Oppenheimer in 1932. Prior to this, any disease of the small intestine was believed to be the result of intestinal tuberculosis. The incidence rates of Crohn's disease have steadily increased to the point that it is now considered a major healthcare problem in both Europe and North America (Hermon-Taylor 2009). By the end of the 20<sup>th</sup> century, it was found in increasing numbers throughout Asia and Australia. This increase is thought to be due in part to the roles played by nutrition and environmental factors in the development of the disease (Barnett et al. 2010). It is currently believed that there is a prevalence of between 1 and 1.5 in 1000 in Western countries (Barrett and Chandra 2011; Rioux et al. 2007). The incidence rates of Crohn's disease have steadily increased to the point that it is now considered a major healthcare problem in both Europe and North America (Hermon-Taylor 2009). In the United States alone, autoimmune diseases affect more than 50 million individuals, which is more than are affected by heart disease and cancer combined (Alexander 2014). It is estimated that the direct cost of care for one patient with Crohn's disease is \$12,417 annually (Feagan et al. 2000). This means that the annual cost of care for all Crohn's disease patients in the United States alone is nearly 8 billion dollars. By the end of the 20<sup>th</sup> century, Crohn's disease was found in increasing numbers throughout Asia, South America, and Australia. This increase is thought to be caused by the nutrition and environmental factors associated with Westernization that are involved in the development of the disease (Barnett et al. 2010; Fofanova et al. 2016). Some scientists believed that this increase in IBD

may not reflect an actual increase in the disease, but rather better access to physician care in areas where the diseases were not previously reported. However, there has also been a significant rise in IBD in Eastern Europe, an area that already had quality healthcare, in the past 25 years. This suggests that the “appropriation of a Western lifestyle” is likely the cause of the rise in IBD, rather than there just being an increase in reported practices (Fofanova et al. 2016).

Crohn’s disease and ulcerative colitis are the two major types of inflammatory bowel disease (IBD), and it is now believed that Crohn’s disease may be one of the most significant risk factors of developing ulcerative colitis (Risques et al. 2006). There are currently around 1,400,000 individuals in the United States with IBD, and there are approximately 30,000 new cases reported every year (Risques et al. 2006). One study has found that as many as 1 in 500 individuals may be suffering from Crohn’s disease, yet many of these have gone undiagnosed (Hubbard and Cadwell 2011).

It is difficult to distinguish between Crohn’s disease and ulcerative colitis, as they share many symptoms and many susceptibility loci. Typically, serological markers are used to diagnose inflammatory bowel diseases (Mathieu et al. 2011). These markers, however, cannot distinguish between the two disease phenotypes. A recent study used gene expression profiling to attempt to differentiate between Crohn’s disease and ulcerative colitis. They used a mathematical ratio on 18 gene expression profiles and found that 37 out of 38 individuals with ulcerative colitis had a score  $> 0$ , while 0 out of 41 Crohn’s disease patients received a score  $> 0$  (Horst et al. 2011). This indicates that the differences between the two diseases may be less related to genetic differences than they are to epigenetic differences that affect the expression of the genes. Another similar disease is celiac disease, which shares fourteen susceptibility loci with Crohn’s disease and is similar in pathophysiology (Cho and Brant 2011).

Migration studies have found that when individuals move from an area of low incidence of Crohn's disease to an area of high incidence that the incidence rate in the migrating population rises to match that of the host population (Hermon-Taylor 2009).

There is a higher prevalence of Crohn's disease among individuals of Eastern European descent, with the highest prevalence being amongst Ashkenazi Jews (Barrett and Chandra 2011; Cho and Brant 2011; Peter et al. 2011). Many of the genes that have been found to be associated with Crohn's disease show population differences between individuals of different descent groups. In particular, many genes that are associated with Crohn's disease patients in European populations (such as *NOD2*, *ATG16L1*, *IRGM*, *5q31*, *IL-23*, and *10q24*) have not been found to be associated with Crohn's disease patients in Asian populations (Cho and Brant 2011; Murdoch et al. 2011). The existence of genetic risk factors specific to the Ashkenazi Jewish population have yet to be determined, but it is believed that determining specific risk factors for specific populations will help lead to more accurate diagnostic and therapeutic treatments (Peter et al. 2011). The prevalence of Crohn's disease amongst Caucasian Canadians is at least twelve times greater than that amongst the First Nations people of Manitoba living in the same environment (Murdoch et al. 2011). These population differences should be taken into account in any study that attempts to understand the epidemiology of the disease.

Crohn's disease affects females more often than males (Barrett and Chandra 2011). The mean age of diagnosis of Crohn's disease is between 33 and 45 years of age (Barrett and Chandra 2011). It should be noted that the age of diagnosis is not necessarily the age of onset of the disease. The author of this dissertation developed Crohn's disease at age 18, but was undiagnosed until age 25. While it does happen, it is rare for the age of onset of Crohn's disease to occur prior to the second decade of life (Hubbard and Cadwell 2011). According to the

Crohn's and Colitis Foundation of America, the mean age of diagnosis for Crohn's disease is between 15-30 years of age (2016).

The genetic mutations that are associated with Crohn's disease are found primarily in industrial societies. This fact leads to some interesting anthropological questions. What is it about such societies that would lead to these mutations? What environmental risk factors are found in industrial societies that are not found in other societies? How has living in a modern society and being exposed to a different diet, different stresses, and increased antibiotic usage led to an increased prevalence of Crohn's disease? How is selection operating on the genes in question? Why is natural selection operating on the gene(s) – what force is driving the selection? These are the questions that will be addressed throughout this dissertation.

One way in which to address some of these questions is by looking at the epigenetic methylation profiles of individuals with Crohn's disease. As a person ages, the amount of DNA methylation found in the genome increases. Monozygotic twins have been found to have methylation profiles that diverge as they grow older (Backdahl 2010; Barres and Zierath 2011). These different methylation profiles may help to explain the discordance of Crohn's disease phenotype that has been found in multiple twin studies.

## ***AUTOIMMUNE DISEASES***

An autoimmune disease is one where the immune system fails to tolerate either the tissues within its own body or the commensal flora within its own intestines. There are over 100 different autoimmune diseases, and they affect a variety of organs and biological systems throughout the body. Autoimmune diseases affect approximately 5-10% of the world's general population (Invernizzi et al. 2009). In the United States alone, autoimmune diseases affect more than 50 million individuals, which is more than are affected by heart disease and cancer combined (Alexander 2014). Autoimmune diseases “result from a failure to control autoreactive immune cells, and a number of anomalies in immune regulatory pathways have been characterized” (Invernizzi and Gershwin 2009). There are numerous autoimmune diseases, including (but not limited to): autoimmune thyroid disease, celiac disease, Crohn's disease, rheumatoid arthritis, multiple sclerosis, primary biliary cirrhosis, psoriasis, systemic lupus erythematosus, and type I diabetes. Most autoimmune diseases have no cure, and direct healthcare costs attributed to these diseases in the United States are estimated to be more than \$100 billion a year (American Autoimmune Related Diseases Association 2014).

The study of autoimmune diseases is fairly recent in human history. While the first reported reference to immunity occurred in 430 BC during what is known as the “Plague of Athens” (Retief and Cilliers 1998) it was not until the 19<sup>th</sup> century that the greatest advancements in immunology occurred: viruses were confirmed as human pathogens, the understanding of humoral (antibody-mediated) and cellular (cell-mediated) immunity arose, and scientists learned of the specificity of antibody-antigen interactions (Metchnikoff 1905; Plotkin 2005). To date, the cause of the Plague of Athens is unknown, though scientists hypothesize that it was caused by either the bubonic plague, smallpox, typhus, toxic shock syndrome, or Ebola (Langmuir 1985;



Olson et al. 1996). The first reported case of a malfunction in the immune system was in 1819 (Velasquez-Manoff 2012). Dr. John Bostock, an English physician, presented a case to the Medical and Chirurgical Society of London in which he described an affliction to the eyes and chest that occurred periodically in mid-June. We now know this affliction as hay fever. He believed that it was caused by the sun or heat, and it was not until fifty years later that another physician (Dr. Charles Blackley) discovered that pollen was the culprit. He noted that those individuals most exposed to the pollen (farmers) were the least likely to develop hay fever (he posited that continual exposure to the pollen made one immune) and he predicted that continued urbanization would increase the prevalence of the disorder (Velasquez-Manoff 2012). His prediction came true – not just in the prevalence of hay fever, but also in the prevalence of all autoimmune diseases. In epidemiological terms, prevalence is the percentage of individuals within a population that have been found to have a condition. The condition is typically either a disease or a risk factor. Incidence is a measure of the probability of a disease occurring within a specific time frame. Incidence is sometimes used to state the number of new cases of a disease that occur within a specific time.

The immune system is designed to maintain a balance in response to pathogens, while at the same time tolerating self-antigens (Alexander 2014). In some individuals, this balance is disrupted: the immune system loses self-tolerance and fights off the very cells it should be protecting. This is what happens in the case of autoimmune diseases. According to the National Institute of Allergy and Infectious Diseases at the National Institute of Health, “in patients with an autoimmune disorder, the immune system can’t tell the difference between healthy body tissue and antigens. The result is an immune response that destroys normal body tissues” (NIH 2008). These diseases have a variety of phenotypes, but they are all similar in that they are

caused by a breakdown somewhere in the immune system. This paper will explore the normal functioning of an immune system and how a person's own immune system can cause an autoimmune disease in general, and Crohn's disease in particular.

### ***WHY STUDY AUTOIMMUNE DISEASES?***

Autoimmune diseases are multifactorial, having environmental and genetic components that determine both the onset and the progression of the disease (Invernizzi and Gershwin 2009). Most autoimmune diseases are polygenic, meaning that multiple genetic mutations contribute to the onset, severity, and progression of the disease. Of the more than 100 known autoimmune diseases, the specific cause has only been identified for 15 of them (Alexander 2014). For the rest, only circumstantial evidence is available at this time.

The prevalence of autoimmune diseases in Western society is on the rise, with the diseases being almost three times as common now as they were only forty years ago (Nakazawa 2008). In fact, autoimmune diseases are on the rise in every industrial nation in the world. This has caused some scientists to label it "the Western disease" (Nakazawa 2008). The increase in prevalence in autoimmune diseases is different for each disorder (see Figure 1). Since the mid-20<sup>th</sup> century, there has been a fourfold increase in individuals with celiac disease and Crohn's disease, a doubling of the number of individuals with asthma, a nearly threefold increase in individuals with multiple sclerosis, and a greater than threefold increase in individuals with type I diabetes, which is expected to double again by the year 2020 (Velasquez-Manoff 2012).

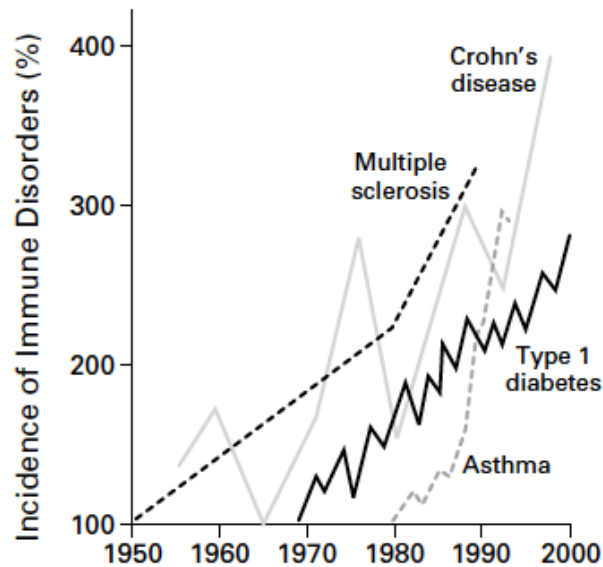


Figure 1 Incidence of Immune Disorders for four autoimmune diseases. (Bach). The x-axis denotes the years 1950-2000. The y-axis denotes the percent increase of immune disorders. Crohn's disease has the largest increase. The data for this graph was obtained from four separate articles (Dubois et al. 1998; Pugliatti et al. 2001; Swarbrick et al. 2001; Tuomilehto et al. 1999).

Scientists believe that it is not a coincidence that there is an increase in autoimmune disease prevalence at the same time there is a decrease in infectious diseases (Velasquez-Manoff 2012). As more and more diseases are being eradicated by vaccinations, there has been an increase in autoimmune diseases. In Sardinia, prevalence of autoimmune diseases began drastically increasing in the 1950s just after the eradication of malaria. “Malaria may have selected for autoimmunity-prone genes. But infection with the malaria parasite *Plasmodium falciparum* likely protected against the dark side of the very genes it helped shape” (Velasquez-Manoff 2012). Once individuals were no longer being infected with *Plasmodium falciparum*, they were left vulnerable to autoimmune diseases. Therefore, as humans try to make themselves healthier by eradicating certain diseases, our immune systems have not adapted to the newer, pathogen-sparse environment, which is leading to more and more cases of autoimmune diseases, and thus unhealthier humans in the long run.

The prevalence of newly diagnosed autoimmune diseases is also rising dramatically among children, as is the prevalence of other syndromes in which the autoimmune system becomes hypersensitive, such as food allergies and asthma (Nakazawa 2008). These increased rates in children are also thought to be primarily due to environmental conditions acting on genetic susceptibility. This increase in autoimmune diseases is believed by scientists to not be due to increased recognition of the diseases or more accurate diagnostic criteria, but rather to something in our environment that is creating increased susceptibility to autoimmune diseases (Nakazawa 2008). The percentage of individuals who have an autoimmune disease is much higher in urban areas than it is in rural areas. The increasing prevalence of autoimmune diseases in industrialized countries, along with the higher frequency of the disease in urban areas, lends credence to the idea that there is a man-made environmental factor or factors contributing heavily to the autoimmune disease phenotype.

#### ***GENETIC CONTRIBUTION TO AUTOIMMUNE DISEASES***

Although the specific genes for many autoimmune diseases have yet to be discovered, there is a great deal of evidence that there is a large genetic contribution to these diseases (Abbas et al. 2014; Adams and Adams 2013; Alexander 2014; Barrett et al. 2008; Barrett and Chandra 2011; Binder and Orholm 1996; Bonen and Cho 2003; Cheon 2013; Cho 2001; Di Sabatino et al. 2013; Hu and Peter 2013; Invernizzi and Gershwin 2009; Mahmoudi 2014; Nakazawa 2008; Ng et al. 2012; Pena 2006; Petronis and Petroniene 2000; Satsangi 1998; Sompayrac 2012; Wagner et al. 2011; Waterman et al. 2011; Zheng 2003). While each autoimmune disease may have a genetic component, the degree to which that component contributes to an autoimmune phenotype varies for each disease.

It has been observed that different autoimmune diseases often exist within a single family. This implies that there is a pleiotropic effect occurring, in which a common gene or genes influences multiple autoimmune disease phenotypes. In fact, it has been shown that mutations in a single gene can lead to multiple autoimmune disease phenotypes. For example, mutations in the *PTPN22* gene, which codes for a protein that affects the responsiveness of T and B cells, are associated with autoimmune thyroid disease, type I diabetes, and rheumatoid arthritis (Invernizzi and Gershwin 2009). It should be noted that these same mutations confer protection against Crohn's disease, another autoimmune disease. Therefore, mutations in a gene can predispose an individual for one type of autoimmune disease while at the same time protect that individual from another autoimmune disease. One example of this is the T300A mutation in the *ATG16L1* gene, which is associated with Crohn's disease, yet confers a protective effect in ulcerative colitis (Serbati et al. 2014).

Also, many different autoimmune diseases share common etiological pathways. For example, histamine expression is upregulated in both inflammatory bowel diseases and allergic diseases, yet the two types of diseases otherwise present with vastly different phenotypes (Kotlyar et al. 2014). Many autoimmune diseases present similar symptoms, such as increased inflammation, although the site of the symptom can vary.

In addition, the prevalence of autoimmune diseases varies between geographic areas. This could be due to either different genetic or different environmental contributions to the disease phenotypes. The frequency of autoimmune diseases by region and ethnicity are only beginning to emerge. To date, autoimmune diseases have primarily been studied within the context of a single disease. Scientists have studied differential susceptibility to that disease, yet few scientists have looked at the susceptibility to autoimmune diseases as a whole within a

region or ethnic group. What is known, however, is that while autoimmune diseases were the 8<sup>th</sup> leading cause of death in females under age 65 in 1995 in the United States, these same diseases were the 6<sup>th</sup> leading cause of death in the United Kingdom with a higher prevalence rate found in that region (Alexander 2014). It is believed that the prevalence of autoimmune diseases is lower in less developed countries, but exact figures are not known at this time.

Type I diabetes is one disease in which the incidence is well documented in different geographical regions. Finland has the highest incidence of type I diabetes, and there is a 350-fold greater incidence rate there than in China, which has the lowest known rates of type I diabetes (Velasquez-Manoff 2012). While some of these differences in autoimmune disease prevalence may be due to genetics, much of it is due to environmental conditions. When individuals move from a low-risk area to an area at high risk of autoimmune disease, their children have the same incidence of autoimmune disease as others living in that region, if not higher rates (Velasquez-Manoff 2012). This shows that the environment plays a huge role in the development of autoimmune diseases. It should also be noted that this difference in incidence of type I diabetes may also be due to differences in the health care systems of the two countries. It is possible that there are more cases of type I diabetes present in China than what has been shown, but that individuals with the disease are untreated, undiagnosed, or underreported.

Also, monozygotic twins have a higher concordance rate of both having a disease than dizygotic twins. The concordance rate is the probability that a pair of individuals will both have a specific characteristic (or disease), given that one of the pair has that characteristic. In other words, if one twin has an autoimmune disease, the odds are much greater that a monozygotic twin will also have the disease than that a dizygotic twin will have the disease. It should be noted that the concordance rate differs between autoimmune diseases (see Table 1), as well as within

autoimmune diseases, and in no case does the concordance reach 100%, which indicates that factors other than genetics must also be involved in the pathogenesis of the disease (Alexander 2014; Cooper et al. 1999). The concordance rates of monozygotic twins and dizygotic twins come from twin registries in many different countries, explaining the differences within autoimmune diseases.

*Table 1 Autoimmune concordance rates in twins*

Disease	Monozygotic Concordance rates	Dizygotic Concordance rates
Rheumatoid arthritis	12% - 30%	3 - 10%
Systemic lupus	25% - 75%	1% - 5%
Multiple sclerosis	20% - 35%	3% - 10%
Type I Diabetes	30% - 50%	5% - 15%
Crohn's disease	50% - 70%	0% - 10%
Celiac disease	70% - 95%	10% - 20%
Graves' disease	40% - 80%	0% - 5%
Ulcerative colitis	5% - 20%	0% - 3%

Concordance rates for eight autoimmune diseases in monozygotic twins and in dizygotic twins (Alexander 2014; Cooper et al. 1999; Greco et al. 2002; Tysk et al. 1988). The data comes from twin registries in North America, Denmark, Finland, Sweden, Japan, the United Kingdom, France, and Australia.

Finally, there is a high rate of correlation between autoimmune diseases and major histocompatibility complex (Consortium) polymorphisms, as well as between autoimmune diseases and non-MHC polymorphisms. For example, individuals with the DR3 and DR4 human leukocyte antigens (HLA) have a much higher risk of developing Type I Diabetes than individuals without those specific HLA alleles (Mahmoudi 2014). The HLA system is a gene complex that encodes the MHC proteins in humans and is responsible for the regulation of the human immune system. The HLA gene complex is located on chromosome 6p21 and contains 3 Mbp (million base pairs). The HLA genes are highly polymorphic, having many different alleles.

The major histocompatibility complex is the set of molecules encoded by a family of genes that control the majority of the immune system. The MHC is encoded on chromosome 6, and contains 240 genes, of which approximately half have known immune system functions (Consortium 1999). There are three classes of MHC: MHC class I, MHC class II, and MHC class III. HLAs of MHC class I present peptides from inside the cell and attract killer T-cells that destroy the cells. HLAs of MHC class II present antigens from outside the cell to T-lymphocytes in order to ultimately stimulate antibody-producing B-cells that will make antibodies to that specific antigen. The MHC class II consists of six major antigen-presenting proteins: DP, DM, DOA, DOB, DQ, and DR). The HLAs of MHC class III are involved in the complement system, which will be discussed later. The HLA is very diverse, which means that the chance that two individuals will have identical HLA molecules on all loci is very low. This difference in HLA molecules is responsible for most organ transplant rejections.

The major histocompatibility complex is one way in which your cells are supposed to be recognized by your immune system as “self”. The cells in your body display the major histocompatibility complex on them, and this complex is not present on foreign cells (Velasquez-Manoff 2012). By displaying the MHC, your body is telling your immune system that these cells belong to you and are to be left alone; however, sometimes the immune system fails to recognize the MHC and attacks healthy cells. While sometimes it is a good thing for your immune system to attack “self” cells (such as when a cell is not functioning properly and needs to be destroyed), autoimmune diseases occur when the wrong cells are continually attacked and the suppressor cells (lymphocytes that can suppress antibody production of other lymphoid cells) are either absent or are ineffective in reining in the cells performing the attack (Velasquez-Manoff 2012).



In addition to autoimmune diseases being polygenic, the genes are also pleiotropic (Alexander 2014). This means that a single gene is affecting multiple traits, not all of them relating to the disease phenotype. For example, a gene that may confer resistance to malaria may at the same time predispose an individual to an autoimmune disease. Mutations in the *NOD2* gene have been found to confer protection from developing ulcerative colitis, while at the same time predisposing the individual to Crohn's disease (Barrett and Chandra 2011). The *FOXP3* gene codes for a transcription factor that helps with the development and function of regulatory T-cells. A single mutation in this gene can lead to a systemic multi-organ autoimmune disease known as immunodysregulation polyendocrinopathy enteropathy X-linked syndrome, or IPEX (Abbas et al. 2014). IPEX manifests with psoriasis, eczema, nail dystrophy, alopecia, autoimmune endocrinopathies, enlarged lymphoid organs, Type I diabetes, food allergies, and recurrent infections.

Females are more prone to developing autoimmune diseases than males (see Figure 2) at a ratio of up to 10:1 (Invernizzi and Gershwin 2009). According to the National Center for Health Statistics, one in nine women will develop an autoimmune disease, while only one in 20 women will develop cancer in her lifetime. Many autoimmune diseases are debilitating, fatal, and more is spent on them yearly than on cancer treatments, yet 10 times more money is spent each year on cancer research than on autoimmune disease research (Nakazawa 2008). The susceptibility of females to an autoimmune disease is different for each disease. The highest female preponderance is 95% to Hashimoto's thyroiditis, while the lowest is just slightly less than 50% to Type I diabetes and ulcerative colitis (Alexander 2014).

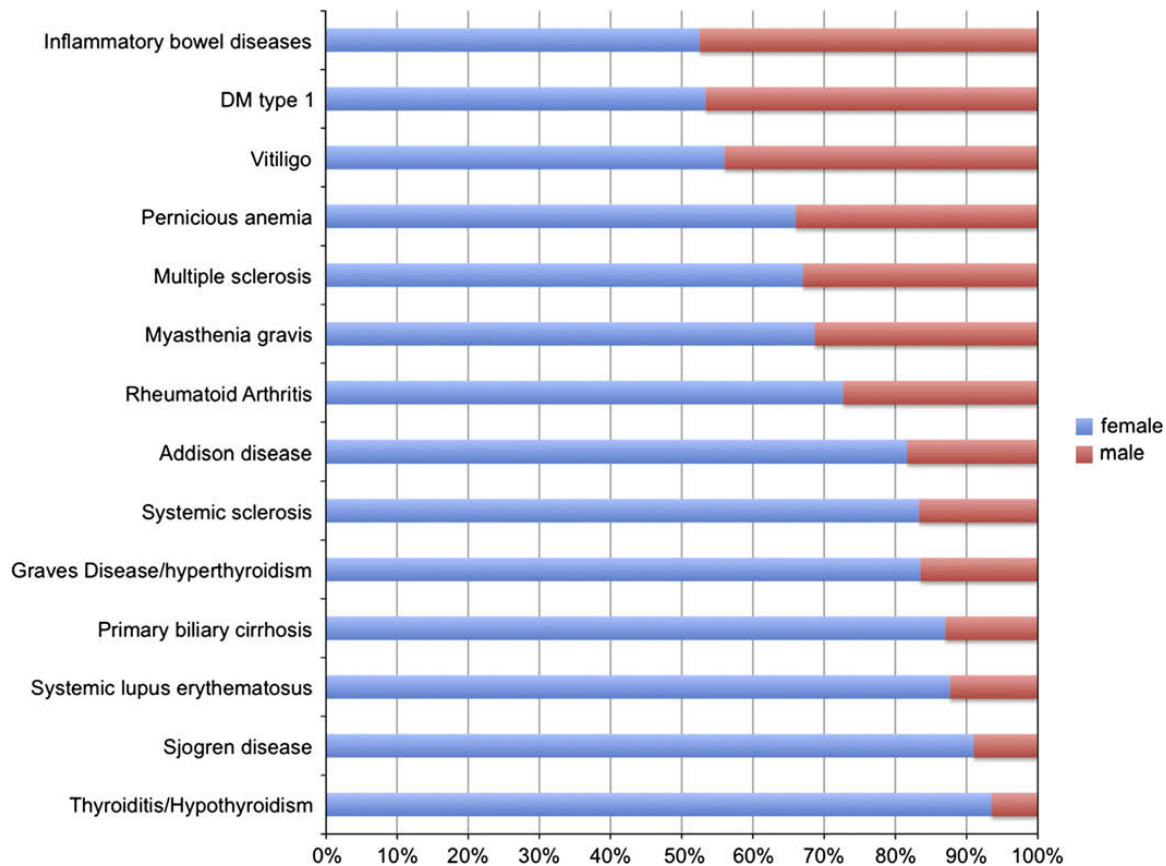


Figure 2 Female to male ratio of selected autoimmune disease expressed as percentage (x axis) and calculated as average (Invernizzi et al. 2009). The populations studied were in the United States and Denmark. Inflammatory bowel diseases include Crohn's disease and Ulcerative colitis.

Besides the HLA genes, many other genes have been found to be associated with autoimmune diseases. It has been suggested that the increased susceptibility of females to autoimmune diseases is due in part to the role of X chromosome gene dosage in autoimmune diseases. Gene dosage is the number of copies of a particular gene within a genome. In some cases (such as in Alzheimer's), only one gene or a few genes exhibit enhanced gene dosage, while in other cases (such as in trisomy 21), an entire chromosome exhibits enhanced gene dosage. In women with autoimmune diseases, these females show evidence of enhanced X chromosome gene dosage through either inactivation or duplication of the X chromosome, indicating that sex chromosomes may be a factor in autoimmune disease susceptibility. Some

autoimmune diseases, such as Type 2 autoimmune hepatitis, present with enhanced gene dosage, while others, such as many thyroid diseases, present with reduced gene dosage (Invernizzi and Gershwin 2009). It has been found that females with autoimmune diseases have a higher rate of circulating cells with a preferentially expressed X chromosome than those females without autoimmune diseases (Invernizzi et al. 2009). Usually a female retains roughly 50% of cells that have the paternal copy of the X chromosome and 50% of cells that have the maternal copy of the X chromosome. In other words, when X chromosome inactivation occurs, the same X chromosome is not always inactivated (Invernizzi et al. 2009). In these women, one of their X chromosomes was preferentially silenced in all of their cells, making them essentially X- rather than XX.

The onset of many autoimmune diseases is during the middle adult years (see Table 2). This is the time of childbearing for women, and thus can have an impact on subsequent populations due to the fact that these women may not be able to reproduce and pass their genes on to the next generation (Alexander 2014). Scientists have debated whether autoimmune diseases are triggered by pregnancy or whether the onset of the disease during childbearing years is coincidence. When a woman gives birth, the cells from the fetus begin circulating in their mothers' blood during the first trimester of pregnancy and can be seen in the bone marrow of women decades later (Choi 2011). Some autoimmune diseases make it difficult or impossible for a woman to carry a fetus to term. According to the American Congress of Obstetricians and Gynecologists (Gynecologists 2013), women with antiphospholipid syndrome, diabetes mellitus, and polycystic ovary syndrome have increased risks of repeated miscarriages and fetal deaths. Other autoimmune diseases (such as lupus nephritis) have medications that must be taken to control the disease that can lead to birth defects, and can make it impossible for women to

breastfeed their offspring. In a 2011 study, the Food and Drug Administration found that 91% of all medications approved for use in adults from 1980 – 2010 lacked sufficient data to determine whether or not they could cause birth defects during pregnancy (Adam et al. 2011). All of this must be considered by a woman with an autoimmune disease to determine whether or not she will be able to be a biological mother.

A longitudinal study in Denmark comparing women who have given birth to those who have not found that women who had a vaginal delivery were 15% more likely to develop an autoimmune disease, women who had a cesarean section were 30% more likely to develop an autoimmune disease, and women who had an abortion were 30% *less* likely to develop an autoimmune disease (likely due to an increase in pluripotent cells entering the woman's bloodstream) (Choi 2011). As was stated earlier, autoimmune diseases were the 8<sup>th</sup> leading cause of death in females younger than 65 years of age in 1995 (Alexander 2014). When cause of death statistics are issued each year, autoimmune diseases are not considered a single cause, but rather more than 20 different diseases. It has been estimated by the American Autoimmune association that if autoimmune diseases were considered a single cause of death then it would be one of the top 5 causes of death for women under the age of 65 in the United States. Thus, many women with autoimmune diseases die prior to reproducing, resulting in the loss of the genetic contribution from the population. At this point, scientists are unsure why the rates of autoimmune diseases are increasing, though environmental factors are a likely culprit.

*Table 2 Mean age of diagnosis for six autoimmune diseases.*

Disease	Mean age of diagnosis
Type I Diabetes	10
Crohn's disease	25
Ulcerative colitis	29

Multiple sclerosis	35
Myasthenia gravis	40
Systemic lupus	40

### ***NORMAL FUNCTIONING OF THE IMMUNE SYSTEM***

As has been shown, autoimmune diseases have a variety of phenotypes, but they are all similar in that they are caused by a breakdown somewhere in the immune system. There are numerous places where this breakdown can occur. It could be in the cytokines, the macrophages, the B-cells signaling molecules, the T-cells, the intracellular signaling molecules, the autophages, the transcription factors, or the cell surface receptors (Invernizzi and Gershwin 2009). Each of these areas will be discussed in the following section.

There are two types of immunity: innate immunity and adaptive immunity (Abbas et al. 2014; Mahmoudi 2014). Innate immunity is also known as natural immunity and it is the first line of defense against pathogens. It does not require memory, which means that your body reacts against an antigen the very first time it comes into contact with it. Innate immunity involves a physical barrier, secretions, and phagocytosis (Mahmoudi 2014). The physical barrier is the skin and the mucosal linings of both the respiratory and gastrointestinal tract. These barriers work to prevent the entrance of foreign organisms into the body. The secretions are tears, saliva, sweat, and gastric acid in the stomach, all of which work to breakdown cell walls and/or prevent bacterial colonization. Phagocytosis is the means by which host cells engulf and destroy antigens. There are two types of phagocytic cells: neutrophils and macrophages. Neutrophils are the predominant white cells in the blood that work to break down the cells walls of bacteria following phagocytosis (Mahmoudi 2014).

### *Innate Immunity: Macrophages*

Macrophages are large white blood cells (approximately 21 micrometers in diameter) that engulf and digest antigens through the process of phagocytosis (Karaiskos et al. 2011). They increase inflammation, stimulate the immune system, and can also decrease immune reactions by releasing cytokines. There are different types of macrophages that each has a different immune response. The M1 cells cause inflammation, the M2a cells are involved in repair of tissue, and the M2c cells are regulatory cells that work to limit the amount of inflammation produced by the M1 cells (Karaiskos et al. 2011). Macrophages can express the following proteins on their surface: CD11b, CD14, CD16, CD33, CD64, CD68, chemokine receptor CCR2, scavenger receptor CD163, co-stimulating molecule CD40, mannose receptor CD206, EMR1, lysozyme M, and Mac-1/Mac-3. Individuals with autoimmune diseases often have a defect in the M2 type of macrophage, and thus their immune system is unable to limit and control the amount of inflammation produced by the M1 macrophages (Karaiskos et al. 2011).

### *Innate Immunity: Phagocytosis*

As mentioned earlier, the three parts of innate immunity are physical barriers, secretions, and phagocytosis. One way in which phagocytosis (the mechanism that involves the host cells engulfing and destroying a foreign organism) occurs is through the complement pathway. A complement is a group of 30 or more plasma and cell surface proteins that work to fight antigens through phagocytosis, inflammation and lysis (Mahmoudi 2014). These proteins are made primarily in the liver and are then distributed throughout the blood and tissues of the body. They work together to form complement pathways. There are five steps involved in complement pathways (see Figure 3).

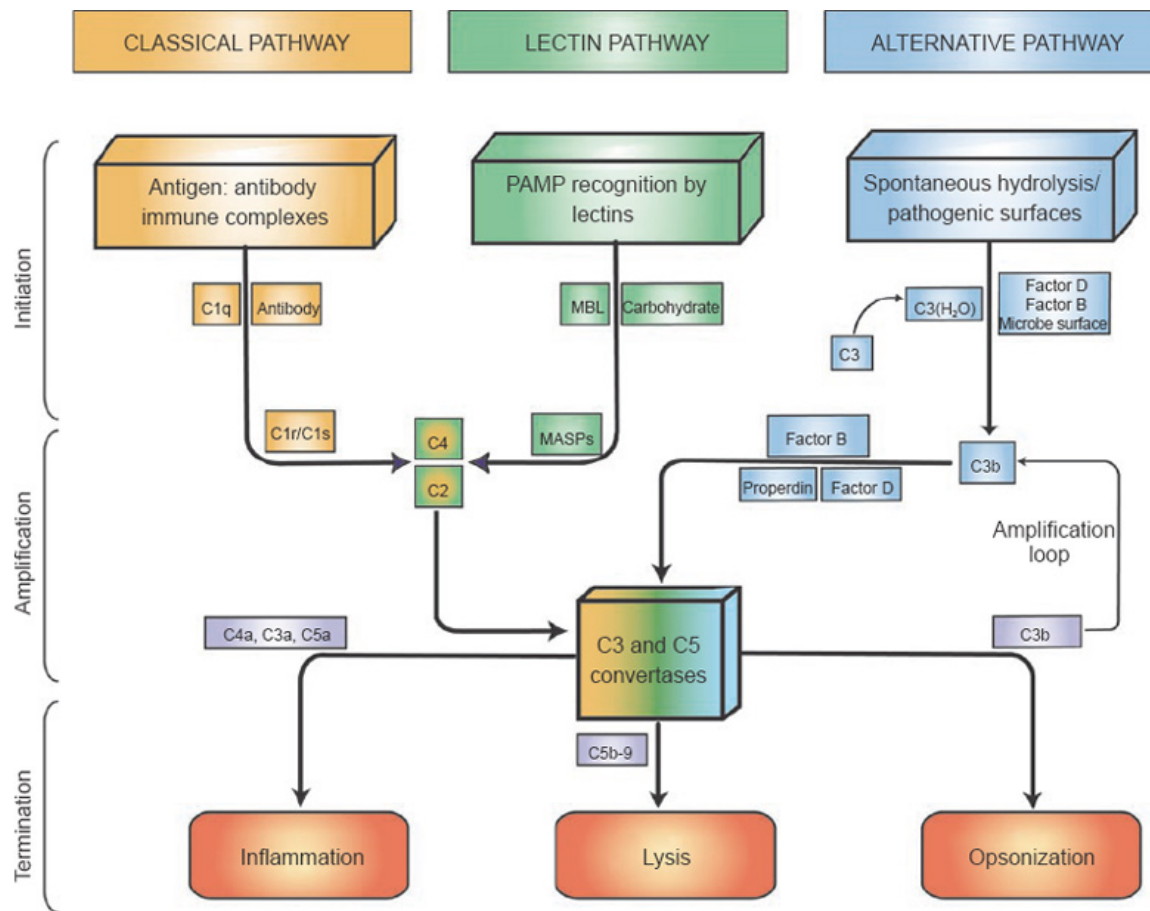


Figure 3 The Complement Pathway (Dunkelberger and Song).

**Step 1** of the complement pathway is activation. This happens through the classical pathway, the alternative pathway, or the lectin pathway. In the classical pathway, the first component (C1) binds to the antigen-antibody complex (Mahmoudi 2014). C1 is comprised of one molecule of C1q (a protein), two molecules of C1r (a serine protease) and two molecules of C1s (another serine protease). This complex is formed when C1q binds to one IgM or six IgG molecules. There are five classes of antibodies: IgA, IgD, IgE, IgG, and IgM. IgM is expressed on the surface of B cells and eliminates pathogens in the early stages of humoral immunity, whereas IgG provides the majority of antibody-based immunity against invading pathogens (Woof and Burton 2004). The other components involved in the classical pathway are C2, C3,

and C4. An antigen is a substance that binds to an antibody or T-cell receptor and elicits an immune response. An antibody is a glycoprotein that is produced by plasma cells in response to an antigen. In the alternative pathway, activation is initiated directly on microbial cell surfaces independent of antibodies. The components of the alternative pathway are properdin, Factor B, Factor D, and C3. In the lectin pathway, activation begins when the mannose binding lectin protein recognizes residues of mannose on the microbial surface and activates the mannose-binding-lectin-associated proteases, MASP-1 and MASP-2 (Mahmoudi 2014). Mannose is a sugar monomer that differs from glucose by inversion of the C-2 chiral center. It is found on the surface of many microorganisms.

**Step 2** of the complement pathway is formation of the enzyme C3 convertase (Mahmoudi 2014). C3 convertase is formed after binding and enzymatic events have occurred. There are two types of C3 convertase, C3bBb and C4b2a. The C3 convertase cleaves C3 into C3a and C3b.

**Step 3** of the complement pathway is opsonization and phagocytosis (Mahmoudi 2014). During opsonization, a microbe is either bound (by an opsonin molecule) or coated by an antibody or a complement. This identifies the microbe as a target for phagocytosis. Phagocytes then bind to the opsonized molecule and engulf it.

**Step 4** of the complement pathway is inflammation (Mahmoudi 2014). C3a and C5a attract neutrophils, as well as bind to mast cells and basophils. Neutrophils are the predominant white cells in the blood. Mast cells are found in the skin and connective tissue. Mast cells bind to Ig-E-coated antigens and then release histamine, heparin chondroitin sulfate, and proteases. Basophils are granulocytes that are found in the peripheral blood and bind to an IgE-coated antigen to release histamine, leukotriene C4 (causes bronchoconstriction), TNF- $\alpha$ , and IL-4. The binding to mast cells and basophils causes degranulation, which releases histamine and other



vasoactive substances. Histamine binds to the endothelial layer of capillaries causing the capillaries to dilate, which then allows fluid and proteins to leak to adjacent tissues, ultimately leading to inflammation (Mahmoudi 2014). This is particularly important in regards to autoimmune diseases. Autoimmune diseases are characterized by the normal immune responses of the body being turned against its own tissues, which results in prolonged inflammation and subsequent tissue destruction.

**Step 5** of the complement pathway is lysis (Mahmoudi 2014). Lysis is the breaking down of a cell through viral, enzymic, or osmotic mechanisms. C3b cleaves C5 to form C5a and C5b. C5b forms a complex with C6, C7, C8, and C9 on the surface of the microbe. This is known as the membrane attack complex. This causes the breakdown of the microbial cell.

#### *Innate Immunity: Natural Killer cells*

Another way in which the innate immunity works is through Natural Killer cells. Natural Killer (NK) cells identify and destroy host cells that have been infected by viruses and bacteria. They make up approximately 10-15% of the lymphocytes in peripheral blood (Mahmoudi 2014). NK cells must be able to destroy infected cells, while at the same time leave non-infected host cells alone. They do this through activating inhibitory receptors. The inhibitory receptors of NK cells bind to the major histocompatibility complex 1 molecule found on the surface of normal cells to prevent the normal host cells from harm. This inhibits the NK cell from activation. When a host cell is infected, the MHC-1 expression is inhibited and has no ligand to which the inhibitory receptor can bind. When this happens, the activating receptors on the NK cells are expressed. The NK cell then binds to and destroys the infected cells. NK cells contain granzyme, which is an enzyme that enters infected cells and causes the cell to undergo apoptosis, also known as programmed cell death (Mahmoudi 2014).

### *Innate Immunity: Cytokine Production*

Natural killer cells produce interferon- $\gamma$ , a cytokine that activates macrophages to destroy microbes that have been phagocytized (Mahmoudi 2014). Cytokines are small proteins that can affect the behavior of other cells through cell signaling. They include chemokines, interferons (IFN), interleukins (IL), colony stimulating factors (CSF), and tumor necrosis factors (TNF). Various cells throughout the body produce chemokines, and these chemokines work to regulate chemotaxis, which is the cellular response to a chemical stimulus. In positive chemotaxis, the cellular response moves towards the chemical in question, while in negative chemotaxis the cellular response moves away from the presence of the chemical. Lymphocytes, dendritic cells, macrophages, fibroblasts, natural killer cells, and T-cells all produce interferons. Interferons activate NK cells and macrophages, and work to increase antigen presentation. Interleukins are produced by leukocytes and cause the stimulation and proliferation of T-cells, B-cells, and NK cells. Colony Stimulating Factors are produced by bone marrow stem cells and by T0 cells. They stimulate the differentiation of progenitor cells into granulocytes, monocytes, and erythrocytes. Tumor Necrosis Factors are produced by macrophages and other cells and work to induce fever, produce proteins, cause apoptosis, and cause septic shock (Mahmoudi 2014).

B-cells, macrophages, mast cells, T-cells, endothelial cells, fibroblasts, and stromal cells can all manufacture cytokines. They work to regulate the maturation, growth, and response of other cells and they act as a balance between humoral and cell-based immune responses. They are usually active in response to infection, inflammation, trauma, sepsis, cancer, and sexual reproduction. There is a strong correlation between cytokine genes and many autoimmune diseases, such as celiac disease, Crohn's disease, rheumatoid arthritis, and many others (Invernizzi and Gershwin 2009). The first cytokine association discovery was that of the

interleukin 23 (*IL-23*) receptor with Crohn's disease, and was found through GWAS (genome wide association studies). It has since been suggested that there is a balance in the activity between *IL-23* and *IL-12* cytokine pathways that is extremely important in pathogenesis of various autoimmune diseases, such as Crohn's disease, primary biliary cirrhosis, psoriasis, and celiac disease (Invernizzi and Gershwin 2009). *IL-12* is a signaling molecule that regulates the activities of the white blood cells responsible for immunity, and it is one of the molecules that is used to discriminate foreign cells from "self".

#### *Adaptive Immunity*

In addition to innate immunity, there is also adaptive immunity. Adaptive immunity occurs when bacteria or other pathogens overcome or bypass innate immunity. Adaptive immunity can distinguish one specific organism from another and develop a memory of each microbe for use in subsequent exposures (Mahmoudi 2014). There are two types of adaptive immunity: humoral immunity and cell-mediated immunity. B-lymphocytes produce antibodies and are involved in humoral immunity. T-lymphocytes destroy microorganisms and are involved in cell-mediated immunity. T-lymphocytes also help B-cells to produce antibodies.

#### *Adaptive Immunity: Humoral Immunity*

In humans, B-cells originate in the fetal liver and from pluripotent stem cells in the bone marrow; mature B-cells are found in the spleen, lymph nodes, and Peyer's patches in the small intestine (Mahmoudi 2014). During maturation, B-cells that recognize and respond to self-antigens go through a process that changes their receptors. This keeps the body's immune system from destroying normal healthy cells. The change in receptors of B-cells from responding to self-antigens to responding to foreign antigens is called receptor editing (Abbas et al. 2014). B-cells produce specific antibodies to fight specific antigens. They are activated after exposure to an

antigen. The antigen binds to the B-lymphocyte surface receptor, which activates the B lymphocyte – sometimes with the help of CD4+ Helper T-cells (Mahmoudi 2014). Helper T-cells are usually required in response to infectious agents, but they are not required in response to most other antigens. The lymphocytes then proliferate and make clones. The clones differentiate and become either antibody-producing cells (these are also known as plasma cells) or they become memory cells. Plasma cells live for a few days or weeks and then die without reproducing. Memory cells, on the other hand, live for years and are able to recognize subsequent attacks by those antigens to which they were originally exposed. If they are later presented with an antigen that they recognize, then they are quickly able to respond by producing new plasma cells to fight off the antigen.

Plasma cells produce antibodies, also known as immunoglobulins (Ig), which are glycoproteins that are produced in response to antigens. There are five types of immunoglobulins: IgG, IgA, IgM, IgE, and IgD (Mahmoudi 2014). IgG is the most abundant antibody, it is prominent in the secondary immune response (re-infection by a known antigen), and it is able to cross the placenta to protect an unborn fetus. IgA is found in bodily secretions and its function is to provide immunity in the mucosal surfaces of the gastrointestinal tract and the respiratory tract. IgM is the largest immunoglobulin and is prominent in the primary (initial) antibody response; it is a membrane-bound antigen receptor on the surface of B-cells. IgE is produced in response to allergic reactions; it binds to mast cells and basophils, causing these cells to release mediators of inflammation. IgD is a membrane-bound antigen receptor on the surface of B-cells. It signals the B-cells to become activated, and can bind to mast cells and basophils to participate in respiratory immune defense.

*Adaptive Immunity: Cell-Mediated Immunity*

Cell-mediated immunity is responsible for dealing with intracellular bacteria, viruses, and any bacteria that is found out of reach of the antibodies from the innate immunity response (Mahmoudi 2014). As stated earlier, T-cells are the primary players in cell-mediated immunity. T-cells only recognize antigens that are in peptide form and are bound to a MHC complex on the surface of antigen-presenting cells (dendritic cells, macrophages, and B-cells). The precursors of T-cells originate in the bone marrow, and then migrate to the thymus to go through the maturation process before being released into the bloodstream. This results in the expression of T-cell receptors and the CD4 and CD8 co-receptors.

During T-cell maturation, the process of negative selection seeks to eliminate all autoreactive T-cells that are capable of reacting to the self-antigen (Mahmoudi 2014). The process of killing cells is known as apoptosis, or programmed cell death. Negative selection affects both self-reactive CD4<sup>+</sup> T-cells and CD8<sup>+</sup> T-cells (Abbas et al. 2014). The process of negative selection is not perfect. There are numerous self-reactive lymphocytes present in all healthy individuals (Abbas et al. 2014). Thus, a defect in negative selection alone may not be a huge contributing factor to autoimmune diseases. Positive selection in regards to T-cells is the process that allows mature T-cells to recognize peptide-MHC complex cells that are not “self”. Occasionally, there will be a mutation in a gene that can affect the ability of T-cells to recognize self-antigens. For example, a defect in the *FAS* gene impairs the apoptosis of autoreactive B-cells and autoreactive T-cells, eventually leading to the disease known as Autoimmune Lymphoproliferative Syndrome (ALS) (Mahmoudi 2014). This is one of the few autoimmune diseases where the exact predisposing genetic factor is known. Some scientists believe that susceptibility genes interfere with self-tolerance and lead to the persistence of an abundance of autoreactive T-cells and B-cells; environmental factors then trigger cell/tissue injury and

inflammation, thus activating the self-reactive lymphocytes to injure the tissues further; the final result is an autoimmune disease (Abbas et al. 2014). This explanation can be used to explain why individuals with the genetic susceptibility to a disease never show the disease phenotype, while also explaining why individuals exposed to the same environmental triggers may have different responses. Neither the gene nor the environment alone can claim sole credit for the onset of an autoimmune disease; rather it is the interaction between the two that is responsible.

Once T-cells mature and leave the thymus, they move into the bloodstream where they recognize and respond to foreign antigens. In most cases, if a mature T-cell is responsive to self-antigens, one of three things happen: anergy, suppression, or deletion (Abbas et al. 2014). These three mechanisms form peripheral T-cell tolerance, which is the ability of the T-cells to respond to foreign antigens while ignoring self-antigens. Anergy is functional inactivation. This occurs when an antigen is recognized, but there is not adequate costimulation. In other words, the T-cells receive signals from their antigen receptors in response to a self-antigen, but they do not receive a signal that there is an innate immune response, which is the second signal required for T-cell proliferation. The anergic cells still survive in the bloodstream, but they are rendered incapable of responding to the self-antigen (Abbas et al. 2014). Suppression occurs via regulatory T-cells. Regulatory T-cells recognize self-antigens and block the activation of lymphocytes that are specific for those antigens. This will be discussed in more detail later. Deletion is programmed cell death – apoptosis. There are two mechanisms by which apoptosis can be induced by self-antigens. In the first mechanism, antigen recognition causes the production of mitochondrial proteins that induce cell death (Abbas et al. 2014). Anti-apoptotic proteins counteract these pro-apoptotic proteins if there is adequate costimulation. In the absence of costimulation, as is seen in the normal response to self-antigens, the anti-apoptotic proteins are

not released and thus cell death occurs. In the second mechanism, recognition of self-antigens may activate the death receptor pathway (Abbas et al. 2014). In the presence of self-antigens, there is co-expression of death receptors and their ligands that activate capsases (“executioner” proteins) to cause apoptosis. For example, the Fas protein is expressed on many cell types, while the *Fas* ligand (*FasL*) is mainly found on activated T-cells. When *FasL* binds to *Fas*, this may induce the death of both T-cells and B-cells that are exposed to self-antigens (Abbas et al. 2014). CD4 and CD8 receptors bind to the antigen-MHC complex at the same time that the T-cell receptors bind to the antigen (Mahmoudi 2014). T-helper cells are known as CD4+ since they have the CD4 co-receptors on their cell surface to interact with the antigen-MHC class II complex. T-cytotoxic cells (killer T-cells) are known as CD8+ since they have CD8 receptors on their cell surface to interact with the antigen-MHC class I complex. MHC class II is present on the antigen-presenting cells: macrophages, B-cells, and dendritic cells. MHC class I is present on all non-red blood cells. Both the CD4 and the CD8 co-receptors send signals from the cell surface into the cell to activate the T-cells.

In order for T-cells to become activated, they must first receive two signals. The first signal is given when the T-cell receptor binds to the MHC-antigen complex of antigen-presenting cells. The second signal is given when the T-cell surface receptor CD28 binds with B7, which is a ligand (a signal-triggering molecule) on the antigen-presenting cell (Mahmoudi 2014). A specific cytokine (*IL-2*) then triggers the T-cells to divide and proliferate. Then, the T-cells begin to differentiate. Some become memory cells, while others become effector cells (CD4+ T helper cells and CD8+ T-cytotoxic cells). CD4+ T helper cells produce cytokines that stimulate macrophages for phagocytosis, prevent viral replication, activate natural killer cells, activate macrophages, activate T-cytotoxic cells to attack MHC I complexes that contain specific

viruses or tumors, stimulate B-cells to proliferate, and stimulate other T-helper cells to proliferate and become T-helper memory cells (Mahmoudi 2014). For the purposes of studying autoimmune diseases, the CD4<sup>+</sup> T helper cells are especially important. It was by studying these cells that scientists were able to first learn the mechanisms of self-tolerance (Abbas et al. 2014). They are also important due to the fact that CD4<sup>+</sup> T helper cells are responsible for orchestrating nearly all of the immune responses to protein antigens. Thus, “failure of tolerance in helper T-cells may result in autoimmunity manifested by T-cell-mediated attack against self-antigens or by the production of autoantibodies against self-proteins” (Abbas et al. 2014). CD8<sup>+</sup> T-cells kill virus-infected cells and tumor cells. Dipeptidyl peptidase IV (also known as CD26) is a serine-type protease that cleaves N-terminal dipeptides away from polypeptides that have either proline or alanine as the next to last amino acid in the sequence (Rose et al. 2002). This protease has been shown to modulate the immune response of helper T-cells. It works by expanding the proliferative response of T-cells and can induce T-cell activation and the production of *IL-2* molecules.

Th cells differentiate themselves from a common CD4<sup>+</sup> progenitor (Th0) in response to stimuli from an antigen or from signals from antigen-presenting cells. The Th1 cells are the ones that activate macrophages. These cells produce interferon gamma (IFN- $\gamma$ ) and are responsible for regulating immune responses to intracellular viruses and bacteria (Backdahl et al. 2009). The Th2 cells produce antibodies and inhibit macrophage production. The differentiation of Th1 and Th2 cells is controlled by histone methylation, histone acetylation and DNA methylation (Backdahl et al. 2009). In order for Th2 to be made, the Th1 has to be silenced through one of the above mechanisms, while Th1 cannot be made without Th2 first being silenced. In the case



of autoimmune diseases, this system breaks down and you no longer have a balance of Th1 and Th2 cells, which is what leads to the chronic inflammation.

The Th2 responses in mammals (including humans) are believed to have originally evolved to confer a resistance to parasites such as helminths (Le Souef et al. 2006). Helminths are very common in the tropical environments in which humans first evolved. When humans moved into cooler and drier climates where helminths are rare, the genetic ability to produce high Th2 responses was no longer needed to fight off those particular parasites. Instead, these high Th2 responses in non-helminth areas led to an increased predisposition to asthma and other allergic diseases. This can be seen occurring in modern populations when individuals migrate from tropical locations to more temperate climates, where many of the migrants develop asthma (Le Souef et al. 2006). Th2 is mediated by IgE, which is the main antibody produced by B-cells activated by *IL-4* (Barrios et al. 2006; Le Souef et al. 2006). It has been found that fewer infections during early childhood favors the development of high levels of Th2 responses (Varner 2002). Th2 responses protect infants and young children against infections and inflammation due to common antigens until their immune system has fully matured.

Regulatory T-cells (Tregs) work to maintain immunological homeostasis, prevent autoimmune responses, and promote self-tolerance (Backdahl et al. 2009). Treg cells are a subset of autoreactive T-cells that help maintain self-tolerance. Treg cells work by inhibiting the functions and activation of effector T-cells, thus making any autoreactive T0 cell unable to attack self-antigens. Approximately 5-10% of all CD4+ T-cells are Treg cells. It has been reported that many common autoimmune diseases (such as Multiple Sclerosis and Diabetes) have down-regulation of the regulatory T-cells (Backdahl et al. 2009). This means that there are fewer Tregs circulating in the bloodstream. Tregs are regulated by epigenetic mechanisms, such as non-

methyated CpGs in the promoter region of the *FOXP3* locus. It has been found that in areas where malaria is endemic, people have a higher concentration of circulating Tregs, while in urban areas individuals have fewer Tregs in their bloodstream (Velasquez-Manoff 2012). The vast majority of malarial cases occur in rural areas (Padilla et al. 2015). It has only been in the past thirty years that malarial cases have begun to be reported in urban areas. A longitudinal study on areas of seasonally endemic malarial transmission found that after times of malaria exposure those living in rural areas had higher levels of circulating Tregs than those living in urban areas (Finney et al. 2009). This could be one of the many reasons why autoimmune diseases are found in increasingly higher frequencies in urban areas than in rural areas. Those individuals in rural areas are more often exposed to factors in our environment (such as malaria-carrying mosquitoes and pollen) that may cause disease, and thus their immune systems function better to withstand such diseases. With a well-functioning immune system that is working hard to fight off diseases not seen in urban areas, there are fewer cases of autoimmune problems. It is only in areas where the immune system does not have to fight off the diseases it has been evolved to fight, as well as being exposed to new man-made environmental factors (such as smog, carbon dioxide emissions, exposure to silica dust, asbestos exposure, tobacco smoke, etc....) to which it has not yet adapted, that you find a high prevalence of autoimmune diseases.

#### *How Autoimmune Diseases Occur*

There are various mechanisms that can explain how the breakdown of self-tolerance leads to autoimmune diseases. One such mechanism is molecular mimicry. Some bacteria have two areas on an antigen that are recognized by an antigen receptor. One receptor resembles the structure of a self-antigen, while the other structure resembles a non-self-antigen (Mahmoudi 2014). The immune system responds to both determinants. An example of this is rheumatic fever.

The bacteria that causes rheumatic fever, *Streptococcus pyogenes*, has determinants on it that mimic the cellular structure of the hosts' heart, heart valves, and nerve cell membranes. When the B-cells produce antibodies against the bacterial infection, they also produce antibodies against the individual's own heart tissues, valves, and nerve cells (Mahmoudi 2014). This can lead to pericarditis, heart failure, and death.

Another mechanism to explain the breakdown of self-tolerance is polyclonal lymphocyte activation. Polyclonal activators, such as lectins (plant proteins), can activate both B-cells and T-cells. These activators are not specific, thus there is a chance of them activating autoreactive cells, which will lead to an autoimmune reaction (Mahmoudi 2014). Examples of polyclonal lymphocyte activation that lead to a disease can be found in the immune response to cytomegalovirus and Epstein-Barr viral infections.

Autoimmune diseases may occur in a specific organ (such as in thyroiditis), or it may involve organs and systems throughout the body (such as in systemic lupus). In general, the following is what occurs in autoimmune diseases. Activation of the T-cell or B-cell in response to a self-antigen causes tissue damage, either by directly binding to cell-surface autoantigens or by forming antibody-antigen complexes that are then deposited in the tissues (Alexander 2014). The tissue damage then leads to the release of cytokines, activated T-cells, and additional self-antigens, which then further stimulate the immune response. The process is cyclic in that the presence of self-antigens triggers the immune system, which triggers the formation of new self-antigens, which further triggers the immune system. The presence of an autoantibody (one that fights against the self-antigen) does not necessarily mean that a person has an autoimmune disease, however. There is some autoimmunity present in all individuals, which indicates that

there are other factors (genetic, epigenetic, or environmental) at work that are involved in the development of an autoimmune disease (Alexander 2014).

A characteristic that is typical of chronic inflammation (which is often seen in many autoimmune diseases) is an influx of activated CD4<sup>+</sup> Th (T helper) cells to the site of the inflammation (Backdahl 2010; Backdahl et al. 2009). Many of these Th cells fail to recognize a cell as “self” and initiate a fight with what they believe to be “foreign” cells by eliciting targeted killing of the cell in question. Normally, there is a balance to the Th cells that have the “kill” response, but in cases of chronic inflammation it has been found that the Th cell lineages that should recognize cells as “self” have been distorted and no longer adequately perform their function (Backdahl et al. 2009).

Another type of helper T-cell, the Th17 cell, has recently been discovered to be involved in the development of chronic inflammatory diseases, such as autoimmune encephalitis (Backdahl et al. 2009). The Th17 cell is a pro-inflammatory cell that secretes *IL-17* and works to mediate inflammation. The differentiation of Th0 cells into Th17 cells is stimulated by Transformation Growth Factor- $\beta$  (*TGF $\beta$* ) and *IL-6* (Backdahl et al. 2009). Th1 and Th2 cells are suppressed by *TGF $\beta$* . Therefore, Th17 is present most often in cases where Th1 and Th2 are either absent or found in diminished amounts. In contrast, the presence of *IL-2* is a growth factor for Th1 and Th2, but at the same time it suppresses differentiation of Th17. *IL-2* is secreted by T-cells when an antigen is present. The fact that Th17 is found in the absence of *IL-2* suggests that Th17 cell differentiation is antigen-independent (Backdahl et al. 2009). The Th17 cells depend on epigenetic modifications, in a way similar to that found in Th1 and Th2 cells (Backdahl et al. 2009). It is typical to find that Th17 cells have histone acetylation. Histone acetylation plays a role in the expression of a gene. It activates transcription and DNA repair (Backdahl et al. 2009).

Individuals with autoimmune diseases generally have abnormalities in their Th cells. Different abnormalities can result in different pathologies in the patient. Crohn's disease and ulcerative colitis are lumped together under the general heading of inflammatory bowel disease in the field of medicine. These two diseases have widely different phenotypes and affect different areas of the gastrointestinal system. The two diseases also differ in the genes that have been found to be associated with them. While there is some overlap in genes that confer risk for both diseases, for the most part different genes affect the onset and severity of each disease. What these diseases do have in common, though, is that they are both affected by abnormalities in Th cells. Individuals with Crohn's disease have a "Th1 inflammatory response involving excessive gamma-interferon production of macrophages" (Rose et al. 2002) that leads to the inflammatory reaction typically seen in Crohn's patients, while ulcerative colitis patients exhibit an abnormal Th2 cytokine profile with an excess production of IL-4 (Rose et al. 2002).

Crohn's disease is a multifactorial autoimmune disease triggered by multiple genetic and environmental factors (Shin et al. 2009). An excellent definition says that it is caused by an "overly aggressive immune response to commensal enteric bacteria in genetically susceptible individuals, where environmental factors precipitate the onset and reactivation of the disease" (Halfvarson 2011). This definition implies that Crohn's disease has a bacterial, environmental, and genetic component.

The first symptoms reported by individuals with Crohn's disease are abdominal pain, bleeding, diarrhea, and weight loss (Barrett and Chandra 2011; Hubbard and Cadwell 2011). Over time, an individual's quality of life is severely affected as the person with the disease will

experience severe diarrhea, severe abdominal pain, fever, weight loss, malnutrition, and bleeding (Dessein et al. 2008).

Crohn's disease is characterized by chronic inflammation of the intestines (Arijs et al. 2011; Hermon-Taylor 2009). Increasing duration of Crohn's disease results in the majority of patients developing complications, including strictures and abscesses/fistulas (Arijs et al. 2011; Murdoch et al. 2011). Circumferential rings of fibrosis (scarring) also develop in many patients with long-term active Crohn's disease (Barrett and Chandra 2011). The scarring and the strictures can each impede the flow of bacteria from one segment of the intestines to another segment, leading to more scarring and fistula formation, and eventually to the entire segment being obstructed and needing to be removed surgically (Barrett and Chandra 2011). The inflammation is often "discontinuous, transmural, and can involve both the small intestine and the colon" (Hubbard and Cadwell 2011)

There are four concentric layers that make up the gastrointestinal tract, including the: mucosa, submucosa, *Muscularis externa* (this is the external muscular layer), and serosa (the smooth membrane of epithelial cells that secretes serous fluid) (Kierszenbaum 2002). Individuals with Crohn's disease have been found to have lymphocytic and granulomatous thickening of all four layers of the intestinal wall (Magin et al. 2013). It has been theorized that this thickening is due to an infectious process of either a viral or bacterial origin that then triggers the genetic predisposition to Crohn's disease (Magin et al. 2013; Sura et al. 2011).

Crohn's disease is often characterized by transmural patchy inflammation from the mouth to the anus (Barrett and Chandra 2011). The inflammation from the disease may affect any portion of the gastrointestinal tract and all layers of the digestive wall (Murdoch et al. 2011). Mucosal expression of interferon gamma (IFN- $\gamma$ ) has been shown to play a role in the

pathogenesis of inflammatory bowel diseases (Gonsky et al. 2014). IFN- $\gamma$  is critical for innate and adaptive immunity against both viral and bacterial infections (Schoenborn and Wilson 2007). It activates macrophages and induces expression of Class II major histocompatibility complex molecules. It is produced by natural killer and natural killer T-cells as part of the innate immune response, and it is produced by CD4 Th1 and CD8 cytotoxic T lymphocyte effector T-cells after the start of cell-mediated immunity (Schoenborn and Wilson 2007). During periods of inflammation, innate immunity cells such as neutrophils and macrophages induce formation of reactive nitrogen species and reactive oxygen species to mediate things such as apoptosis, cell proliferation, cell repair, and immune modulation (Backdahl et al. 2009).

Human polymorphonuclear leukocyte (PMN) elastase is a mediator of inflammation on connective tissue, and there is a strong correlation between the concentration of PMN-elastase and patients with active Crohn's disease (Dhôte et al. 2000). Polymorphonuclear leukocyte elastase is stored in the granules of the PMNs and works to degrade structural and soluble proteins. PMN-elastase acts as a marker for the degranulation activity of leukocytes (Dhôte et al. 2000). High concentrations of PMN-elastase have been found in the plasma of patients with a variety of inflammatory diseases. High levels of PMNs is a typical finding within the mucosa of individuals with inflammatory bowel diseases, and it is indicative of a period of acute inflammation (Dhôte et al. 2000). Since high concentrations of PMN-elastase are only found in individuals suffering from an outbreak of an inflammatory bowel disease, it cannot be used as a diagnostic tool for the disease, but the levels can work to tell the physician the severity of the outbreak.

Children and adolescents with active Crohn's disease tend to have stunted growth, and severe weight loss is found in adults with the disease (Annunziata et al. 2011). These

characteristics of the disease may be due to growth hormone resistance caused by the persistent chronic inflammation that is typically found in individuals with Crohn's disease (Annunziata et al. 2011). The mean age of diagnosis of Crohn's disease is 26, with a range of 9 to 79 years (Halfvarson 2011).

There are many commonalities between Crohn's disease and other autoimmune diseases, such as certain allergic disorders, particularly o-pulmonary allergic diseases. For example, histamine and mast cell activity is similar in both Crohn's disease and allergic disorders (Kotlyar et al. 2014). IgE is a key immunoglobulin involved in both Crohn's disease and allergic disorders. Crohn's disease and many allergic disorders often show benefits from the use of probiotics and nutritional interventions (Kotlyar et al. 2014). By studying the immunological pathways of one (either Crohn's disease or non-pulmonary allergic diseases), it is hopeful that improved treatments will be discovered that will also be beneficial to other diseases.

The chronic inflammation associated with Crohn's disease increases the risk of cancer in the small bowel and in the colon (Barrett and Chandra 2011). It is also associated with increased risk of intestinal fibrosis (Honzawa et al. 2011). Intestinal fibrosis is an excessive deposition of extracellular matrix due to chronic inflammation of the intestines. In more than one-third of patients with Crohn's disease, intestinal fibrosis has been known to cause thickening of the bowel wall, forming strictures that may require surgery (Rieder and Fiocchi 2008). While Crohn's disease is primarily thought of as a disease of the intestines, there are also many extraintestinal manifestations that can occur. These include: arthritis, uveitis (inflammation of the eye), severe skin rashes, growth retardation, and hidradenitis suppurativa (boils on or near sweat glands or sebaceous glands) (Barrett and Chandra 2011). These manifestations occur in over 20% of Crohn's disease patients, and an individual may suffer from any or all of them.



While much is known about autoimmune diseases and Crohn's disease, in particular, there is still much to learn. A better understanding of the malfunctions that occur in the immune system and their causes will aid in the development of more effective treatments for those individuals suffering from an autoimmune disease. This review has shown the way in which a normal immune system works, how that immune system is turned against itself in autoimmune diseases, and the known mechanisms that occur in Crohn's disease. Due to the fact that hundreds of billions of health care dollars are spent in the United States alone each year on individuals with autoimmune diseases, more information on what causes the immune system to malfunction and how to repair it is essential to the overall health of Western and Westernizing society.

## ***BACTERIAL/VIRAL THEORY OF CROHN'S DISEASE***

There have been multiple studies that have found that Crohn's disease could be the result of a bacterial infection, as will be demonstrated below. There are also studies, however, that have concluded that bacterial exposure does not affect whether or not an individual will develop Crohn's disease (Sokol et al. 2010).

Commensal flora are the bacteria that are normally contained within the intestines of an individual. There are over 1,000 types of bacteria normally found in the intestines (Barrett and Chandra 2011). The commensal flora are present from birth, and can be affected by diet, antibiotics, and infection (Barrett and Chandra 2011). Crohn's disease usually involves an aberrant response to the intestinal bacteria normally found within an individual (Hubbard and Cadwell 2011)

One of the environmental factors associated with Crohn's disease is exposure to *Mycobacterium paratuberculosis* (MAP), a subspecies of *Mycobacterium avium*, as this bacteria has been found to cause a similar pathology of intestinal inflammation in animals known as John's disease (Shin et al. 2009).

One study suggests that Crohn's disease is the result of impaired macrophage responses to intracellular pathogens following exposure to pathogens such as *Listeria*, *Salmonella*, and *Mycobacteria* (Behr 2010). Individuals with Crohn's disease do not exhibit inflammation in the mesenteric fat (Batra et al. 2011). It is believed that the reason for this lack of inflammation is due to the presence of macrophages. In particular, there is an increase in macrophage subset 2 (M2) cells in the mesenteric fat, possibly providing a protective role to the fat (Batra et al. 2011).

One study found that “most people with Crohn’s disease” have been exposed to, and possibly infected by, *Mycobacteria avium* (Hermon-Taylor 2009). A later study tested for the presence of the genetic sequence of *Mycobacteria avium* in 81 Crohn’s disease patients and 85 healthy controls, and found that neither DNA nor RNA from *Mycobacterium avium* was present in any of the individuals (Sasikala et al. 2009). Rather than look for the presence of MAP genetics to determine whether or not a person had been exposed to *Mycobacterium*, one study looked at the proliferation of T-cells and the cytokine response (Sibartie et al. 2010). They found that an increase of T-cell proliferation and an altered cytokine response is indicative of some prior exposure to *Mycobacterium avium*, and that individuals with Crohn’s disease exhibit this evidence of exposure to the bacterium. The altered cytokine response that the authors found is that individuals with Crohn’s disease secrete a significantly greater amount of TNF- $\alpha$  in response to MAP stimulation compared to controls. The authors stress that this association does not necessarily imply causation, but rather that exposure to the bacterium may be an “an environmental modifying factor” to the disease (Sibartie et al. 2010).

One study found that there are clusters of Crohn’s disease patients that overlap areas where there is the presence of infectious microorganisms such as MAP in “unpasteurized milk and cheese, untreated water supplied by wells or springs, animal manure used as fertilizer for family vegetable gardens, and bodies of water contaminated by agricultural runoff” (Pierce 2009).

According to Hermon-Taylor (Hermon-Taylor), the result of Crohn’s disease from a bacterial infection is thought to occur in two stages. First, the primary pathogen (particularly *Mycobacterium*) is widely distributed throughout the intestines, causing an immune system response. This immune response can result in the loss of mucosal integrity. Second, secondary

pathogens from the lumen then invade the already compromised wall of the intestines, with the final result being the inflammation that is characteristic of Crohn's disease. These secondary pathogens are typically the normal organisms that live in the intestines, as well as some enteric microbiome. It is defects in the innate immune system that allow the bacteria from the lumen to penetrate through the intestinal epithelial barrier (Barrett and Chandra 2011).

It has been found that individuals with Crohn's disease have a shorter bowel length than individuals without the disease (Walzer and Buchman 2010). This shorter bowel length is prior to any intestinal resection surgery that may be needed. Walzer and Buchman (Walzer and Buchman) propose that this shorter bowel length may predispose people to the development of the disease by altering the intestinal motility and intestinal flora normally found in a normal gut. If the bacterial theory of causation of the disease proves to be true, people with a shorter bowel may be more likely to contract bacterial infections that in turn may lead to the development of Crohn's disease, as well as have abnormal levels of the bacteria normally found in the intestines. To date, six separate studies have found that individuals with Crohn's disease have a significant reduction of *Faecalibacterium prausnitzii* (a commensal gut bacterium) compared to healthy controls (Dave et al. 2011).

Some authors have debated the bacterial theory of the disease. Rather than an aberrant response to bacteria being the cause of the disease, many scientists believe that the bacterial infection is an opportunistic infection that occurs secondary to the onset of Crohn's disease (Hubbard and Cadwell 2011). Part of the reason why this is the belief is that no one species has been found to clearly be a trigger of the disease. Rather, it is possible that the genotype of the individual favors the overgrowth of harmful bacteria in the gastrointestinal tract while blocking the production of the helpful bacteria found there (Hubbard and Cadwell 2011).

Not only can bacterial infections possibly lead to Crohn's disease, but the antibiotics used to treat such infections may also lead to the disease. Taking antibiotics has been found to cause a commensal microbial disruption (Dave et al. 2011; Drouet et al. 2011). Both bacterial infection and antibiotics can cause a short-term disruption of the normal bacterial found in the gut. This short-term disruption alone, however, is not sufficient to lead to Crohn's disease.

In addition to the possible bacterial association with Crohn's disease, it has also been found that some viruses are also found in higher frequencies amongst individuals with Crohn's disease. In 2013, Dimitroulia *et al.* published a study that found there was an association between Crohn's disease and the Epstein-Barr virus. They tested individuals with and without Crohn's disease for the Epstein-Barr virus in both their blood and their tissues. There was not a significant difference between the samples in the blood; however, there were significant differences when looking for the virus in the intestinal tissues. They found that individuals with Crohn's disease were significantly more likely to have the Epstein-Barr virus in their intestines than individuals without the disease (Dimitroulia et al. 2013). They also found that individuals who had active Crohn's disease were significantly more likely to have the virus in their system than those individuals who were in remission from Crohn's disease at the time of the study. This suggests that there is a potential viral involvement not only with the disease itself, but also with the severity of the disease. It is unclear at this point, though, whether the Epstein-Barr virus was causative in the onset of Crohn's disease or whether it is functioning as an opportunistic virus that is better able to survive in the intestines of individuals who are already immunologically compromised due to the Crohn's disease.

Another study also found an association between Crohn's disease and the  $\gamma$ -herpes virus Epstein-Barr virus (Hubbard and Cadwell 2011). This disease is also found in the colonic tissue of patients, which is where Crohn's disease is often found. Therefore, the association between the two may be coincidental, rather than causal. Noroviruses and cytomegalovirus are also often associated with inflammatory bowel diseases (Hubbard and Cadwell 2011; Nguyen et al. 2011). There is no causality link to the cytomegalovirus. The norovirus has been found to exacerbate the symptoms of an individual with an inflammatory bowel disease. In animal model studies, the norovirus is the only disease that has been found to be associated with Crohn's disease thus far (Hubbard and Cadwell 2011). In fact, it has been found that cytomegalovirus is significantly more often associated with ulcerative colitis than it is with Crohn's disease (Nguyen et al. 2011).

As changes in microbiota that are associated with disease pathogenesis are more likely to be evident in new-onset cases, scientists have studied the microbiota of pediatric patients to determine if there are significant associations in Crohn's disease patients as compared to control populations. The microbiota of an individual changes over time, as well as in response to environmental modifications. A study in 2014 retained 45.5 million sequences from 447 pediatric patients and 221 controls and found positive associations between a diagnosis of Crohn's disease and the abundance of *Pasteurellaceae*, *Veillonellaceae*, *Neisseriaceae*, and *Fusobacteriaceae* (Gevers et al. 2014). They also found a negative association between Crohn's disease and the following genera: *Bacteriodes*, *Faecalibacterium*, *Roseburia*, *Blautia*, *Ruminococcus*, *Coprococcus*, *Erysipelotrichaceae*, and *Bifidobacteriaceae*. It is important to note that *Faecalibacterium prausnitzii* is a well-recognized anti-inflammatory organism, the absence of which might help to explain the chronic inflammation in patients with Crohn's disease (Gevers et al. 2014). A more recent study has found that food and alimentary habits

(which are progressively altered in modern societies) contribute to the dysbiosis seen in individuals with IBD compared to phenotypically normal controls (Rapoza et al. 2017).

One recent study looked at the possible association between Crohn's disease and four viruses and four bacteria (Magin et al. 2013). The viruses were the adenovirus, Epstein-Barr virus, herpes simplex virus I, and parvovirus B19. The bacteria tested were *Mycobacterium avium paratuberculosis* (MAP), *Listeria monocytogenes*, *Escherichia coli*, and *Clostridium perfringes*. There was no evidence of adenovirus, Epstein-Barr virus, herpes simplex virus I, parvovirus B19, or MAP in the sampled Crohn's disease individuals (Magin et al. 2013). *Clostridium perfringes*, *Listeria monocytogenes*, and *Escherichia coli* were all found in higher percentages in the controls than they were in individuals with Crohn's disease. This shows that the normal microbiota living in the gastrointestinal tract changes in individuals with Crohn's disease so that they have less of the "good" bacteria than healthy individuals. This is why many individuals with Crohn's disease are highly encouraged to take a probiotic pill daily, and it is also why there have recently been some cases of individuals with Crohn's disease being transplanted with fecal matter from healthy individuals. A meta-analysis on all of the studies on fecal microbiota transplantation through May of 2014 found that 60.5% of individuals entered clinical remission following a transplantation of healthy fecal matter (Colman and Rubin 2014). This analysis included 122 patients with IBD, 39 of whom had Crohn's disease. It should be noted that fecal microbiota transplantation has not been approved for treatment of Crohn's disease by the FDA. A more recent study on nine pediatric Crohn's disease patients found that 5 of 9 patients were in clinical remission of Crohn's disease 12 weeks after transplantation without any additional medications (Suskind et al. 2015). Of those individuals reporting symptoms

following the transplantation, all adverse events were mild, and less than what the patients experienced prior to the transplantation.

It has been found that in order for long-term problems to occur (such as that seen in individuals with Crohn's disease), there must also be a genetic component that predisposes the individual to long-term disruption of the commensal microbial in addition to the bacterial infection, viral infection, and/or antibiotic treatment (Dessein et al. 2008; Drouet et al. 2011; Sura et al. 2011). Drouet *et al.* (Drouet et al.) found evidence that the *NOD2* mutant gene in conjunction with antibiotics induced Crohn's-like symptoms in mice. It is possible that there is a three-way interaction that takes place within individuals who develop Crohn's disease. First, the individual must be genetically predisposed to the disease. Second, the individual is exposed to a viral infection that causes epigenetic changes in his/her genome. Third, the epigenetic changes affect the expression of genes, thus leading to Crohn's disease. This has been shown to be the case in one recent study that found that viral infection may regulate *NOD2* expression in individuals with Crohn's disease, as well as cause an allelic imbalance in the *NOD2* and *ATG16L1* genes (Hu and Peter 2013).



## ***GENETICS OF CROHN'S DISEASE***

Family studies, twin studies, and animal models (Ohmen 1996) have shown that Crohn's disease and ulcerative colitis are both multifactorial diseases with different environmental and genetic components (Binder and Orholm 1996; Hubbard and Cadwell 2011; Satsangi 1998; Zouali 2003). Approximately fifteen percent of Crohn's disease patients have a family member who also has the disease, and there is a thirty-six percent concordance rate amongst monozygotic twins versus a four percent concordance amongst dizygotic twins (Barrett and Chandra 2011; Cho and Brant 2011). A recent study of 68 pairs of Swedish-born twins found that there was not a significant difference between monozygotic and dizygotic twins in terms of incidence rates (Halfvarson 2011). These statistics show that there is definitely a genetic component to the disease, but they also show that there must also be something other than genetics at work. It should be noted that while families are more likely to share disease-associated alleles, they are also more likely to share similar intestinal microbiomes (Cho and Brant 2011). If the basis of Crohn's disease is more bacterial in theory than genetic, the similar microbiotics living in the intestines of family members might show up as a genetic association to the disease when in actuality it is an association to the types of flora that are welcome in the gut of the families in question.

There have been over 100 genes that have been found to be associated with Crohn's Disease (Costello et al. 2005). In some cases, the association appears to be the result of a single nucleotide polymorphism (SNP) in the gene; in some cases, it is a microsatellite; in some cases, it is differential expression of the gene. When a gene is highly methylated, the expression of the gene is usually reduced or completely silenced, but methylation can also cause activation of that gene (Backdahl 2010). The mechanisms by which this occurs are not completely understood at

present. It is believed, however, that it likely involves the recruitment of methyl-CpG binding domain proteins and histone deacetylases (Dabritz and Menheniott 2014). Together the proteins and deacetylases can repress transcription of the gene through the induction of heterochromatin. It should be noted that there is some debate on whether the methylation causes the transcription repression or whether it is a by-product of the process (Dabritz and Menheniott 2014).

According to a 1996 article, the following examples are indicative of the genetic basis of IBD: monozygotic twins have a higher concordance than dizygotic twins; it is rare for there to be an IBD concordance between spouses (who share a similar microbiome); there is a difference in disease frequency between geographic areas and ethnic groups; and it has been shown that there are many instances of family members exhibiting IBD at different times (Ohmen 1996). Of the two, Crohn's disease is the most severe disease, affecting a larger proportion of the population. In the United States alone, 0.2-0.5% of people suffer from Crohn's disease. Some of the environmental components associated with Crohn's disease are diet, smoking habits, parasitic infection, and the influence of hormonal status and drugs (Bridger et al. 2002; Karlinger et al. 2000). The information in Table 3 originates from the article by Fofanova *et al.* (Fofanova et al. 2016) and summarizes the environmental and nutritional factors that are associated with Crohn's disease, as well as their effect on the microbiome and the epigenome.

*Table 3 Environmental and nutritional factors that affect Crohn's disease.*

<b>Factor</b>	<b>Crohn's disease Effect</b>	<b>Microbiome Effect</b>	<b>Epigenome Effect</b>
<b>Helminth infection</b>	Protects against CD	Increases diversity	Affects methylation signature in host T-cell response
<b>Antibiotic use</b>	Increased risk of CD	Decreases diversity	Transgenerational effects on sperm viability
<b>Urbanization</b>	Increased risk of CD	Increase or decrease diversity	Currently unknown

<b>Smoking</b>	Increased risk of CD	Decreases diversity	Global, persistent methylation changes
<b>Preterm birth</b>	Increased risk of CD	Decreases diversity	Affected methylation profile that persists into adulthood
<b>Gastroenteritis</b>	Increased risk of CD	Promotes infection with and replication of enteric viruses	Currently unknown
<b>Mycobacterial infection</b>	Increased risk of CD	Currently unknown	Inhibits chromatin remodeling
<b>Breastfeeding</b>	Protects against CD	Promotes microbial homeostasis	Modifies DNA methylation and histone modification
<b>High fat/high carb diet</b>	Protects against CD	Increases diversity	Affects metabolic pathways through methylation
<b>Dietary fiber</b>	Ameliorates CD	Increases diversity	Modifies DNA methylation and histone modification
<b>Dietary protein</b>	Increases risk of CD	Decreases anti-inflammatory producers	Currently unknown
<b>Vitamin D</b>	Increases risk of CD	Regulates microbiome; increases <i>C. difficile</i> infection risk (common in CD patients)	Lack of vitamin D epigenetically reduces immune cell development
<b>Food additives</b>	Increases risk of CD	Decreases diversity and function	Modifies epigenome to confer colitis susceptibility
<b>Refrigeration</b>	Increases risk of CD	Currently unknown	Currently unknown

Through a combination of linkage analysis, population association studies, positional cloning, and candidate genes, scientists believed in 2005 that there are nine regions on eight separate chromosomes that are associated with IBD. Genes on the nine regions will be examined, followed by recent genes that have been found to be associated with IBD. These regions are on chromosomes 16 (IBD1 and IBD8), 12 (IBD2), 6 (IBD3), 14 (IBD4), 5 (IBD5), 19 (IBD6), 1 (IBD7), and 3 (IBD9). Even though scientists now know that there are more than nine susceptibility regions, the old method of naming IBD1 through IBD9 is still in effect. Each IBD

loci has different genes and different variants that are believed to be associated with Crohn's disease. Recent studies have shown that there are more regions that affect IBD (Barrett et al. 2008). In fact, genome-wide association studies (GWAS) have found that there are over 100 genes that are associated with Crohn's disease (Cho and Brant 2011). One genome-wide association study found 195 genes associated with Crohn's disease (Dinu et al. 2012). This was a very large study involving 22,000 individuals with Crohn's disease and 25,000 controls. It should be noted that all of those genes combined do not explain more than 20% of the total genetic variance seen with Crohn's disease (Rosenstiel et al. 2011). The only chromosomes so far that do not contain risk genes are chromosomes 20, X, and Y (Barrett and Chandra 2011). A recent study looked at 100 genes that have been found to be associated with Crohn's disease and found that the expression of 67 of these genes was significantly different in individuals with active Crohn's disease compared to controls (Palmieri et al. 2011). Approximately half of the genes showed increased expression, while the other half of the genes showed decreased expression. Differential expression of the genes while maintaining identical genotypes, as was found in that study, implies that something other than genetics may also be affecting disease phenotype.

The genes associated with Crohn's disease have a wide variety of functions. It has been discovered that genes with the following functions are all associated with the disease: microbe recognition, lymphocyte activation, cytokine signaling, and intestinal epithelial defense (Cho and Brant 2011). There is differential expression in the functional genes that are associated with Crohn's disease. A recent study found 57 genes that are differentially expressed in Crohn's disease that are not differentially expressed in ulcerative colitis (Lin et al. 2014). Figure 4 illustrates the functionality of these Crohn's specific differentially expressed genes. The genes

that have been predominantly found to be associated with Crohn's disease are those that function to regulate autophagy (Cho and Brant 2011). Autophagy plays a large role in the innate immune response to a variety of intracellular pathogens by encapsulating cytosolic cellular components and then causing their degradation and ultimate recycling into components usable by the cell (Kabi et al. 2012). In Crohn's disease, this pathway is defective, and has led to the hypothesis that the disease may result from an impaired antibacterial response that in turn leads to chronic inflammation.

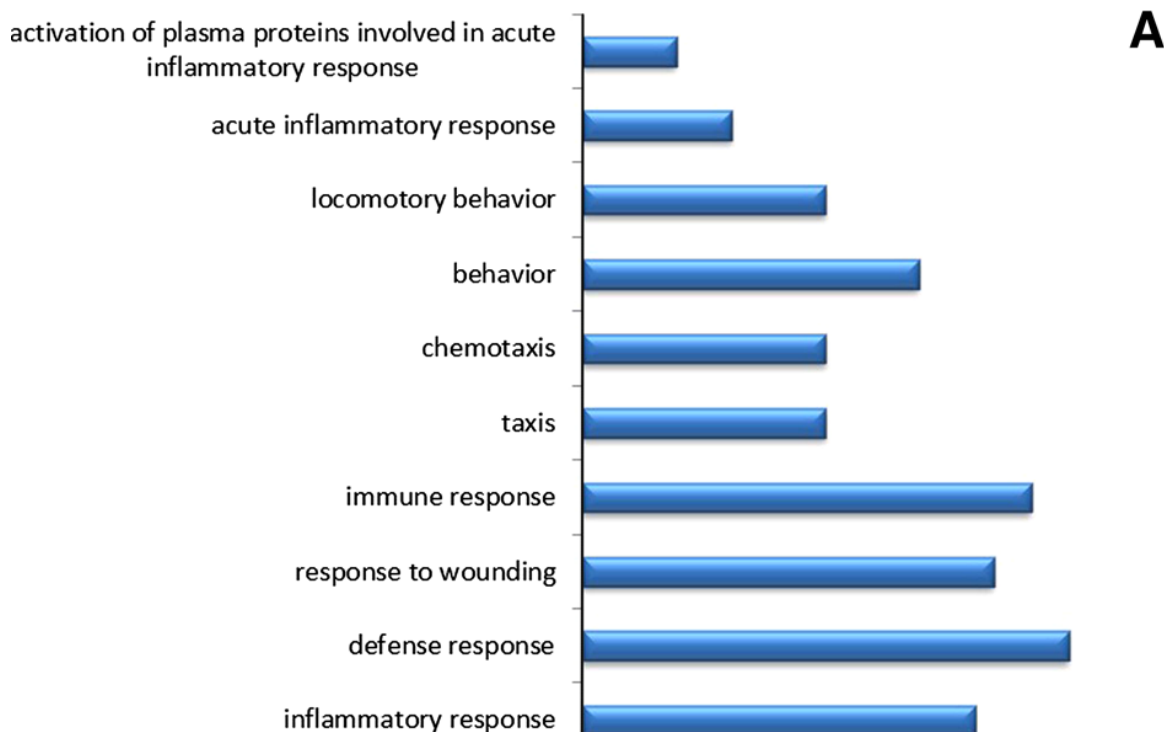


Figure 4 Gene Ontology functional enrichment analysis of differentially expressed genes in Crohn's disease. The x-axis is the number of genes that are differentially expressed. The y-axis shows the increase in p value from bottom to top (Lin et al. 2014).

While some studies found associations between Crohn's disease and these regions, other studies were unable to reproduce the associations. This is common when searching for the genetic basis of multifactorial diseases: much of the research conducted is unreplicable by other scientists. In the case of Crohn's disease, however, there is an additional problem with

replicating research. Due to the similarities between Crohn's disease and ulcerative colitis, many patients are misdiagnosed for years. An initial misdiagnosis of Crohn's disease as ulcerative colitis or vice versa can cause patients to be grouped with the wrong disease population, which can ultimately skew the results. Another possible problem is that many of the studies that have been conducted were conducted on pediatric patients. The problem with this is that Crohn's disease usually affects a person during their 20's and 30's, so some of the people in the "control" group might develop Crohn's disease later in life.

## **IBD1**

Ohmen et al. (Ohmen) found that there is a 26 cM region at IBD1 that is inherited greater than by chance amongst sib pairs (linkage analysis). In this region on chromosome 16 is the nucleotide oligomerization domain (*NOD2*)/Caspase-activation recruitment domains 15 (*CARD15*) gene on 16p12.3, extending to 16q13. It was believed in 2007 that between 25-30% of the genetic susceptibility to Crohn's disease can be explained by mutations in this region (Rodriguez-Bores 2007). More recently, however, scientists have found that all of the susceptibility loci from every chromosome combined only account for 23.2% of the heritability of Crohn's disease (Barrett and Chandra 2011; Nimmo et al. 2012). There are over 30 polymorphisms of the *NOD2/CARD15* gene, with three common variants (Arg702Trp, Gly908Arg, and Leu1007insC) that account for approximately 82% of the mutated alleles (Cho 2001). It should be noted that the *NOD2* mutations associated with Crohn's disease among individuals of European ancestry have not been observed in individuals of either Asian or sub-Saharan African descent (Cho and Brant 2011; Kim et al. 2011). Individuals with complicated Crohn's disease (structuring and/or non-perianal penetrating) are significantly more likely to

have at least one *NOD2* variant than those individuals without the disease – 40% in affected individuals versus 31% in controls (Markowitz et al. 2011).

The first gene found to be associated with Crohn's disease was the *NOD2* gene. This gene consists of two caspase recruitment domains (*CARDs*), a nucleotide binding domain, and a leucine rich repeat domain, which is where the most common *NOD2* mutations are located (Barrett and Chandra 2011). The *NOD2* gene is a pattern recognition receptor of the innate immune system, and the protein this gene makes is essential in distinguishing normal intestinal flora from pathogenic bacteria (Barnett et al. 2010; Murdoch et al. 2011). The *NOD2* gene is also involved in activating many other genes through signaling pathways, such as *IKKY*, *MAVS*, *IRF3*, *IL-10*, etc.... (Kaser and Blumberg 2011; Kim et al. 2011). For the complete signaling pathways, see the drawing in Figure 5 that was in the article by Kaser and Blumberg (Kaser and Blumberg 2011). Mutations in this gene can result in a decrease in intracellular bacterial sensing (Barrett and Chandra 2011). The *NOD2* protein also plays a role in apoptosis, nuclear factor kappa B activation, and the development of an inflammatory response in the epithelial cells (Murdoch et al. 2011). Due to the many different functions provided for by the *NOD2*, as well as the different genes that the *NOD2* gene is responsible for activating, it should come as no surprise that the same *NOD2* mutation that can increase susceptibility to Crohn's disease can at the same time protect the individual from infection by other common enteric bacterium (Kim et al. 2011).

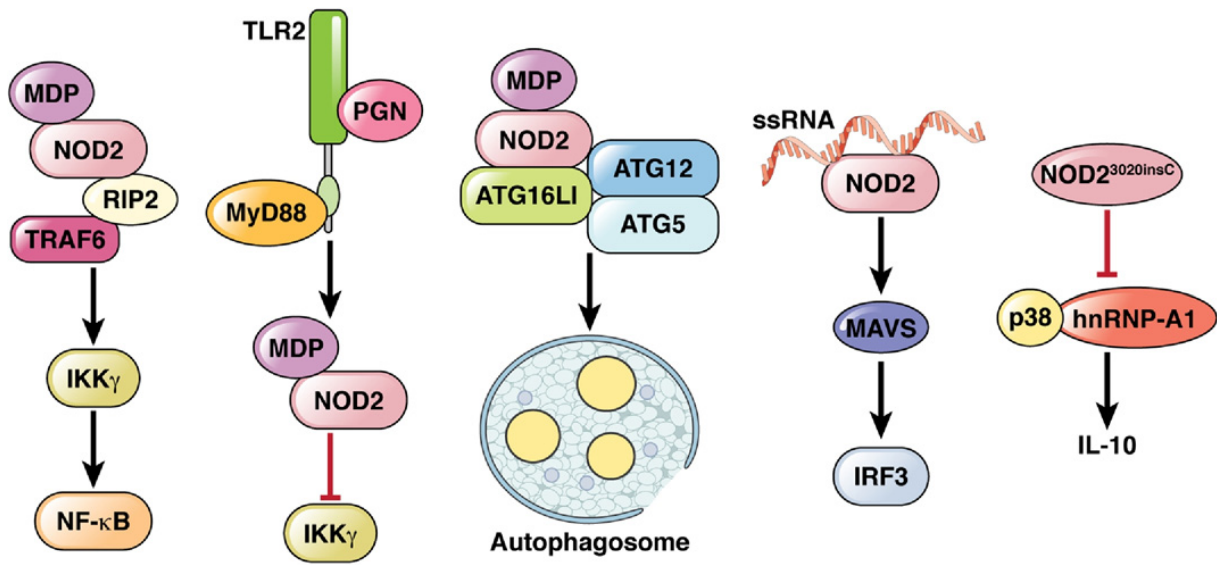


Figure 5 *NOD2* signaling pathways (Kaser and Blumberg).

According to one study, the *NOD2* gene is only expressed in myelomonocytic cells (such as macrophages and granulocytes), dendritic cells, and intestinal epithelial cells (Barrett and Chandra 2011). According to other studies, however, the *NOD2* gene is expressed in antigen presenting cells, macrophages, lymphocytes, Paneth cells, fibroblasts, and epithelial cells (Cho and Brant 2011; Nimmo et al. 2011). Paneth cells are important in the innate immunity system. They are necessary for maintaining the intestinal stem cell niche, for secreting antimicrobial peptides, and are responsible for packaging up invading microbial cells into granules that can be secreted (Hubbard and Cadwell 2011; Kaser and Blumberg 2011). Crohn's disease patients that are homozygous for the ATG16L1 risk allele have abnormal Paneth cells.

The *NOD2* gene is activated by viral ssRNA, which then triggers phosphorylation of transcription factor interferon response factor-3 (IRF3). Translocation of the IRF3 leads to the activation of interferon genes, such as IFN- $\beta$ , which mediates innate immune responses to viral agents (Barrett and Chandra 2011). In other words, in order to properly fight off a viral infection,



the *NOD2* gene must be functioning properly as it is the first step in the pathway that results in an antiviral response. In response to bacterial products, *NOD2* is activated to in turn activate *NF- $\kappa$ B* (nuclear factor kappa-light-chain enhancer of activated B-cells), which is a regulator of inflammation (Barrett and Chandra 2011). *NOD2* activates *NF- $\kappa$ B* in the presence of MDP (muramyl dipeptide) (Hubbard and Cadwell 2011). MDP is a part of bacterial walls that have inflammatory properties. This activation of *NOD2* is controlled by the leucine rich repeat domain, and mutations in this region result in the inability of cells to be able to recognize muramyl dipeptide, a fragment of peptidoglycan found in the cell walls of both gram-positive and gram-negative bacteria (Barrett and Chandra 2011). If the muramyl dipeptide is not recognized by the leucine rich repeat domain of the *NOD2* gene, the body will not make an immune response to the presence of invading bacteria.

The *CARD* gene codes for a protein that is partially responsible for resistance to microbial pathogens, and it also causes a deficit in NF-kappa B activation in response to bacterial agents (Bonen and Cho 2003). The *CARD* gene is involved in innate immunity (Lee and Buchman 2009). Patients with mutations in the *CARD15* gene are unable to control intestinal inflammation (Pena 2006). Forabosco *et al.* found through linkage analysis that when a person had two mutant alleles in this gene the results were a more severe form of Crohn's disease than when a person only had one mutant allele (Forabosco et al. 2000).

Another study that showed the association between Crohn's disease and IBD1 (Farmer et al. 2001) claimed to be performing a population association study, however their "control" population all had some form of gastrointestinal disease, which makes their results suspect. They hypothesized that the use of microsatellite markers in this region may be able to be used as a screening tool to differentiate between Crohn's disease and ulcerative colitis. Bonen and Cho

(Bonen and Cho) found that one mutated copy of the *NOD2/CARD15* carries a 2-4-fold risk to the patient, while two mutated alleles carries a 20-40-fold risk. The *NOD2* gene has been shown to have a somewhat recessive trait quality to it, regardless of which *CARD* allele is with it (Zheng 2003), and can lead to apoptosis (cell death). Individuals with two mutated copies of the *NOD2* gene also tend to exhibit symptoms at a younger age than those individuals who are heterozygotes for the gene (Barrett and Chandra 2011).

In addition to the presence of SNPs in the *NOD2* gene, there is also evidence of differential expression of parts of the gene in individuals with identical genotypes at that loci (Nimmo et al. 2011). A study by Nimmo (Nimmo et al.) found that the expression of *GALNT2* and *vimentin* (two proteins made by the *NOD2* gene) was significantly different in individuals with Crohn's disease as opposed to those without the disease, regardless of the genotype of the individual. This differential expression of the gene lends credence to the theory that there is an epigenetic factor at work in individuals with Crohn's disease.

#### *Natural Selection and the NOD2 locus*

Many alleles associated with complex diseases are subsets of genetic variants that are spread over geographically separated populations (Myles et al. 2008). Most of the Crohn's disease susceptibility alleles that have been identified are found in European populations, but significant associations have not been found in non-European or non-North American populations (Nakagome et al. 2010; Zaahl et al. 2005). This suggests that there is population-specific susceptibility to Crohn's disease. Uneven distributions of particular alleles and haplotypes, as is seen in Crohn's disease susceptibility genes, are thought to be signals of a recent adaptation to the environment through natural selection (Oota et al. 2004). The most well-known example of a disease-causative allele being maintained by environmental adaptation is the

allele that causes sickle cell anemia. It is more difficult to detect polygenic adaptation and positive selection on preexisting (i.e. standing or previously neutral) variation than it is on novel variation (Nakagome et al. 2012). Positive selection on a novel variant leaves a large homogeneous region surrounding the advantageous allele, whereas positive selection on a standing variant is more likely to exist on multiple haplotypes (Innan and Kim 2004).

A recent study examined all of the Crohn's disease associated SNP sites to determine whether or not any of the alleles have undergone positive selection on a previously neutral variant. To do this, one must first find the Crohn's disease risk SNP sites that are shared among all the populations studied. See Table 4 for a list of the populations studied and the number of individuals in the study. The scientists chose SNPs whose expected heterozygosity was estimated to be greater than 0.2 in at least 75% of the populations studied as the SNPs shared amongst all of the populations at intermediate frequencies are likely to predate the Out-of-Africa migration (Nakagome et al. 2012). The SNP sites chosen had spatial intervals of 2-16kb, as linkage disequilibrium blocks typically extend up to 50kb in non-African populations and less than 10kb in African populations due to bottlenecks (Kidd et al. 2004). The ancestral alleles were inferred for all the SNP sites by comparing the human SNPs to that of chimpanzee and orangutan sequences. The authors used a variety of methods to determine that the Crohn's disease haplotype in the *NOD2* locus showed the clearest evidence of having undergone positive selection on a standing variant (Nakagome et al. 2012).

*Table 4 The continent and populations studied in the research by Nakagome et al. (Nakagome et al.), as well as the sample size of each population used in the study.*

Continent	Population	Sample Size
Africa	Biaka Pygmy	70
Africa	Hausa	38
Africa	Mbuti Pygmy	39
Africa	Chagga	45
Europe	Adygei	54

Europe	Irish	112
Europe	Danes	49
East Asia	Russians	48
East Asia	Japanese	95

The authors of the previously mentioned study determined the number of population-specific mutations in the *NOD2* gene for the African, European, and East Asian populations. They then converted those numbers to a ratio of European to African-specific mutations and European to East-Asian-specific mutations. The authors then determined maximum likelihood estimates for the mutation rate and the time to the most recent common ancestor (see Figure 6), as well as ages of mutations for each of the European haplogroups H1, H2, and H3 (see Figure 7). Their results showed that there was a positive selection at the *NOD2* locus present on multiple haplotypes, with the H1 haplotype, which has an increase in the advantageous allele, experiencing positive selection in European populations (Nakagome et al. 2012). The Crohn's disease risk alleles define the root and internal branch of the H1 haplogroup, are found in low frequencies in African populations, and are found independently on other haplotypes. This suggests to the authors that the Crohn's disease risk alleles on *NOD2* locus were present prior to the Out-of-Africa migration, but have undergone selective forces in European populations following the migration (Nakagome et al. 2012). Other loci associated with Crohn's disease have also been shown to experience a selective sweep (Huff et al. 2002).

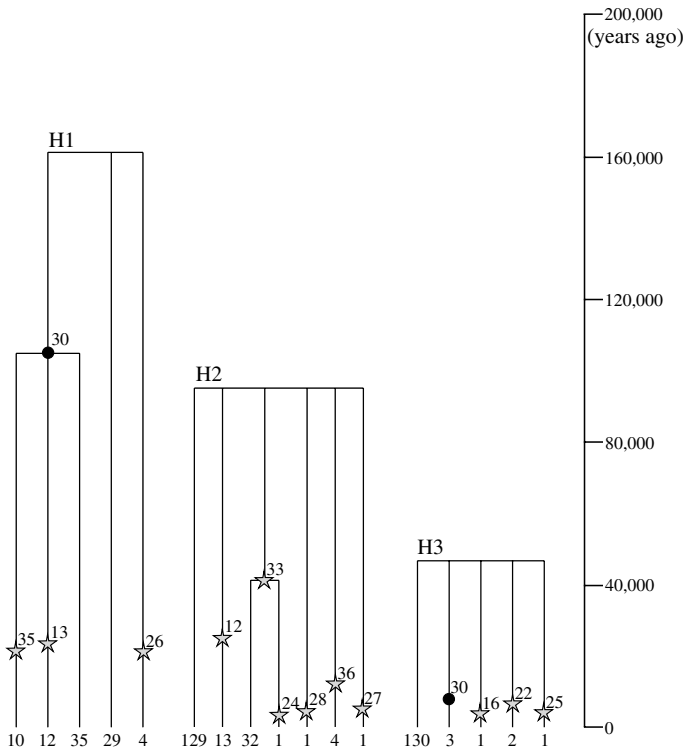


Figure 6 Gene trees for H1, H2, and H3 haplogroups in Europeans. The heights of the trees are proportional to the time to the most common recent ancestor. Black circles indicate mutations shared among Africans and Europeans, whereas stars indicate European-specific mutations. Numbers on the black circles or stars correspond to the SNP# (Nakagome et al. 2012).

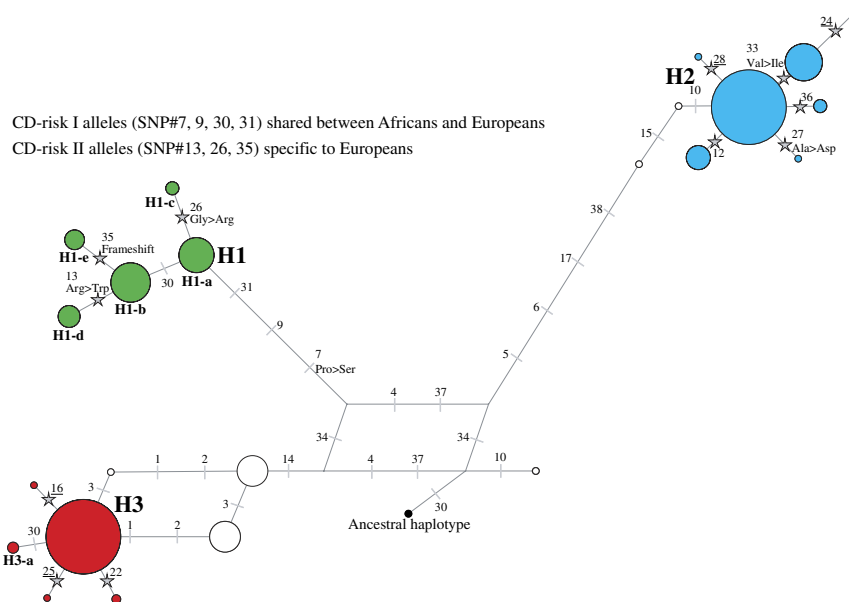


Figure 7 Haplotype network in Europeans based on 38 SNP sites. The size of the circle corresponds to the frequency of the haplotype. The numbers on the branches correspond to the SNP#. The stars on the branches indicate European-specific alleles. Underlined SNP site numbers indicate the sites that the phase inference failed to assign to one of two chromosomes in an individual (Nakagome et al. 2012).

## IBD2

There is a single polymorphism *TaqI* (which is at a Vitamin D receptor) on chromosome 12 that has been found to be associated with Crohn's disease susceptibility by use of linkage analysis (Simmons 2000). Simmons *et al.* also found this allele to be associated with susceptibility to a variety of infectious diseases. A population association study on this region found that there is no significant association between Crohn's disease and *Interleukin-1*, which is where *TaqI* is located (Heresbach et al. 1997). Satsangi *et al.* also found there to be no association between *TaqI* and Crohn's disease (Satsangi 1998). Also in this region and associated with Crohn's disease is *IFN-γ*, which is thought to be related to the reduction of tissue repair in Crohn's disease patients (Leeb et al. 2003).

## IBD3

In this region on chromosome 6 are the major histocompatibility complex genes. It has been found that *HLA-B\*1501* is associated with fistulous Crohn's disease through linkage analysis (Rodriquez-Bores 2007). Satsangi *et al.* performed non-parametric linkage analysis as well as a population association study and found that there was no linkage between Crohn's disease and the HLA gene (Satsangi 1996). One of the major differences between Crohn's disease and ulcerative colitis is that ulcerative colitis is more highly associated with the HLA region than Crohn's disease (Cho and Brant 2011). Many studies, however, have found that there is an association between Crohn's disease and different areas of the HLA complex. Crohn's disease has been found to be associated with many different HLA genes, all of them involved in encoding for class II major histocompatibility complex molecules: *HLA-DRB1\*0103*, *HLA-*

*DRB1\*04*, *HLA-DR7*, *HLA-DRB3\*0301*, *HLA-B18*, *HLA-B21*, *HLA-DR6*, *HLA-DR8*, and *HLA-DR10* (Barrett and Chandra 2011). The HLA genes are important in the regulation of the body's immune response and still might contribute to the severity of Crohn's disease (Pena 2006), even if they are not significantly associated with the initiation and propagation of the disease. Instead, they may be associated with the specific tissues that exhibit Crohn's disease and the severity of the disease (Barrett and Chandra 2011). The association between Crohn's disease and HLA Class II genes indicates involvement of T-cells in the pathology of the disease due to the fact that T-cell receptors on the surface of the cells of the mucosal lining of the intestinal tract recognize and bind HLA molecules (Backdahl et al. 2009).

#### **IBD4**

On this region on chromosome 14q11-12 is *Interleukin-25 (IL-25)* that is associated with Crohn's disease (Rodriquez-Bores 2007). *IL-25* is a proinflammatory cytokine, and mutations in this gene weaken immunity to gastrointestinal infection and chronic intestinal inflammation. This gene was found through a candidate gene approach, and tested through linkage analysis (Rodriquez-Bores 2007).

#### **IBD5**

There are multiple genes on this region on chromosome 5q31-33 that are associated with Crohn's disease: *OCTN1*, *OCTN2*, *DLG5*, *MDR1*, and *IL-6* (Rodriquez-Bores 2007). This region is a 250-kb haplotype that consists of a cytokine gene cluster that is believed to be a general risk factor for both forms of IBD (Giallourakis et al. 2003).

The *OCTN* genes are carnitine/organic cation transporters that harbor a cytokine cluster, and have been shown to be associated with Crohn's disease through population association studies (Török 2005). These genes work to maintain the integrity of the epithelial barrier (Barrett and Chandra 2011). The *OCTN1* and *OCTN2* variants are single nucleotide polymorphisms, but they have been shown to not be independent of the entire IBD5 region through transmission disequilibrium testing (Russell 2006). Mutations in *OCTN1* and *OCTN2* may affect the transporter function of *OCTN1* by “reducing uptake of important biologic compounds (e.g. neurotransmitters) while increasing uptake of toxins, such as putrescine, derived from bacterial catabolism” (Barrett and Chandra 2011).

The *MDR1* gene is a multi-drug resistant drug that produces a protein that reduces the intracellular concentrations of toxins and xenobiotics. It has been found that mutations of this gene cause a significant reduction in response to steroid therapy in patients with Crohn's disease (Boder et al. 2005). This entire region has been found to play a role in the initiation of Crohn's disease (Zheng 2003).

The IBD5 region is responsible for producing the tumor necrosis factor-alpha (*TNF- $\alpha$* ) protein, as well as the interleukin 1 (*IL-1*) and interleukin 6 (*IL-6*) proteins. In patients with Crohn's disease, *TNF- $\alpha$*  has increased secretion, possibly leading to the initiation and propagation of the disease (Sykora 2006). *IL-1* and *IL-6* proteins are associated with proinflammatory cytokines and inflammatory cell infiltration, which are unbalanced in patients with Crohn's disease. It has recently been found that genetic variations in this region can result in impaired barrier function of the intestines (Barrett and Chandra 2011).

## **IBD6**



A linkage peak in a genome-wide scan on chromosome 19 identified this region as a possible genetic contribution to Crohn's disease (Vermiere 2005). A subsequent study (although it was published prior to the previous study) found that there was a significant association between Crohn's disease and patients who were homozygous for an intracellular adhesion molecule 1 (*ICAM-1*) in this region (Low 2004).

### **IBD7**

One of the genes on this region of chromosome 7 is transforming growth factor beta 2 (*TGF-beta 2*), which is a cytokine that plays a role in prevention of autoimmunity and in anti-inflammatory responses. This association was found to be significant through population association testing. It is suggested that this gene may be related to the development of intestinal strictures in Crohn's disease patients (Fell et al. 2000).

### **IBD8**

There is evidence of a second chromosome 16 locus (independent of *NOD2*) that is associated with Crohn's disease, however no studies have been performed to find this locus as of yet (Rodriguez-Bores 2007). In 2011, it was found that haptoglobin allele 2, located on chromosome 8, is significantly associated with both Crohn's disease and ulcerative colitis (Marquez et al. 2011). The normal version of haptoglobin plays a protective role in inflammation by inhibiting the production of Th1 and Th17 cytokines (Marquez et al. 2011), so it makes sense that a mutated form of that allele would not provide inflammatory protection, and thus be associated with Crohn's disease.

## IBD9

The CC-chemokine receptor 5 (*CCR5*) is located on chromosome 3p21. This gene is responsible for regulating T-cells. There has been a statistically significant association found between homozygosity for the *CCR5delta32* variant of this gene and the presence of anal lesions in Crohn's disease patients (Satsangi 1996). *CCR5* is a co-receptor that works with CD4. Many people who are resistant to HIV have the *CCR5delta32* mutation.

A cDNA microarray was used to identify novel unknown genes associated with Crohn's disease (Costello et al. 2005). This study found that there were 81 genes that were differentially regulated (upregulated) between normal controls and patients with Crohn's disease. These genes fell into three main categories: (1) immune and inflammatory response; oncogenesis, cell proliferation, and growth; and (3) structure and permeability. Many of these genes were the ones mentioned above; however, there are dozens of genes that have yet to be studied regarding their role in patients suffering from Crohn's disease. In addition to the above-mentioned chromosomes, Costello *et al.* (Costello et al.) also found association between Crohn's disease and chromosomes 2, 4, 8, 15, 17, and 20. All of these studies combined show that Crohn's disease is a multifactorial disease with many different genetic bases, and it will probably be a long time before the exact etiology of the disease is fully known.

## ATG16L1

The *ATG16L1* gene encodes a 44-kD protein on chromosome 2 that is involved in the intracellular autophagy complex (Fabio et al. 2011b; Kaser and Blumberg 2011). The gene is expressed in intestinal epithelial cells (Rioux et al. 2007). Mutations in this gene are thought to diminish bacterial clearance by impairing the autophagy ability of the cell (Barrett and Chandra

2011; Dessein et al. 2008; Hubbard and Cadwell 2011). Specifically, mutations have been shown to impair the clearance of *Salmonella typhimurium*, one of the most common forms of *Salmonella enterica enterica* (Rioux et al. 2007). In order to understand what this means, it is important to first get a good definition of what autophagy is. “Autophagy degrades damaged organelles and proteins, in homeostasis and as a response to starvation, and is important for the clearance of pathogens (xenophagy), which is required for immunity to multiple different types of bacteria” (Cho and Brant 2011).

Autophagy is necessary in the mediation of the inflammatory response. It works by capturing organelles and cytoplasm and holding them within a membrane-bound organelle known as an autophagosome; once held, a lysosome comes along and destroys the microorganism within the autophagosome through the process xenophagy (Barrett and Chandra 2011). If this process is disrupted, the result is a production of inflammatory cytokines, which may lead to the poor response to intracellular bacteria that is found in individuals with Crohn’s disease (Barrett and Chandra 2011). It has been found that several genetic variants associated with an increased risk of Crohn’s disease are associated with autophagy (Hubbard and Cadwell 2011; Newman et al. 2009).

One study found that the *ATG16L1* risk variant was associated with an increase in *E. coli* proliferation in the body (Elliott et al. 2011). A mutation in the *ATG16L1* gene could also result in the invading pathogens (such as *Salmonella*) not being held within the autophagosome, and thus not destroyed by the lysosome. Specifically, “*ATG16L1* mutations cause disordered production of key anti-microbial peptides made in Paneth cells in the lining of the intestine” (Barrett and Chandra 2011). The odds ratio of this gene in association with Crohn’s disease depends on the number of copies of the Thr300Ala A→G variant the individual has. The odds

ratio for the Thr/Thr genotype is 0.76 in a Mediterranean population; for Thr/Ala genotype, the odds ratio is 0.89; for the Ala/Ala genotype, the odds ratio is 1.42 that the individual will develop Crohn's disease (Fabio et al. 2011b). In a European population, it was found that having one copy of the T300A(+) variant gives an odds ratio of 1.26 (OR of 1.00 for T300A(-) variant) (Fabio et al. 2011b). These results show that the odds ratio of this gene (as well as most other genes) is entirely dependent on the population under study.

It has been found that there is a link between the *NOD2* and the *ATG16L1* genes (Kaser and Blumberg 2011). The activation of *NOD2* via MDP induces *ATG16L1*-dependent autophagy in cells that do not express the risk allele of either of these two genes (Hubbard and Cadwell 2011). This interaction limits intracellular bacterial replication, and it also increases the major histocompatibility (Consortium) class II surface expression (Hubbard and Cadwell 2011)

## **BPI**

The bactericidal/permeability increasing protein (*BPI*) is located on chromosome 20 and is involved in eliminating gram negative bacteria (Klein et al. 2005). There is a common SNP on the *BPI* gene (Glu216Lys). The Glu/Glu genotype is found less frequently in Crohn's disease individuals than in controls, while the Glu/Lys or Lys/Lys genotype is found more frequently in individuals with Crohn's disease than it is in individuals without the disease (Klein et al. 2005). This mutation could be related to the associations found between Crohn's disease and bacterial infections.

## **CCL**

Both Crohn's disease and ulcerative colitis have been found to be associated with genes that are involved in regulating cytokine expression. Both diseases are associated with *chemokine C-C motif ligand 2* (*CCL2*) and 7 (*CCL7*), located in the 17q12 region (Cho and Brant 2011). The *NOD2* gene is responsible for the production of *CCL2* by the stromal cells (Hubbard and Cadwell 2011). Only Crohn's disease, however, is also associated with *chemokine C-C motif receptor 6* (*CCR6*) (Cho and Brant 2011). This gene is expressed in immature dendritic cells and memory T-cells, and functions to mediate leukocyte recruitment during times of inflammation (Cho and Brant 2011).

## **CYLD**

A recent study has found a gene located 9kb downstream of *NOD2* on chromosome 16 that is associated with Crohn's disease (Cleynen et al. 2011). In this gene, there are two SNPs associated with increasing risk of developing the disease, and two SNPs associated with protecting one from developing the disease. *CYLD* is a key negative regulator of *NF-kB*, and it is downregulated in the intestines of individuals with Crohn's disease (Cleynen et al. 2011).

## **FOX03**

The *FOX03* gene is a transcription factor protein located on chromosome 6 that has been shown to be a tumor suppressor that regulates gene expression (Snoeks et al. 2009).

## **FUT2**

Individuals with the W134X allele in their *FUT2* gene do not express ABO blood group antigens in their saliva, and are known as non-secretors (McGovern et al. 2010). Individuals who are homozygous for the W134X allele are protected from infection of some norovirus strains due to their non-secretor status (Hubbard and Cadwell 2011). If these individuals are not susceptible to infection, and if the bacterial theory of Crohn's disease were correct, it would follow that non-secretor individuals would also be less susceptible to developing Crohn's disease. This has been found to be correct (Hubbard and Cadwell 2011).

## **IL-7**

Interleukin gene 7 is a hematopoietic growth factor found in red marrow and the thymus, as well as produced by epithelial and other types of cells. It has been found to be involved in the association between Crohn's disease and intestinal fibrosis by upregulating the expression of *HSP47* (heat shock protein 47) in intestinal fibroblasts (Honzawa et al. 2011).

## **IL-10**

Another interleukin gene found to be associated with Crohn's disease is the *IL-10* gene (Cho and Brant 2011). This gene functions to inhibit the expression of proinflammatory cytokines and increase the expression of anti-inflammatory cytokines. Individuals with a mutated *IL-10* that results in loss of *IL-10* function show "a severe CD phenotype without any apparent environmental trigger" (Cho and Brant 2011).

## **IL12B**

*IL12B* is located on chromosome 5, and is associated with autoimmunity – when *IL12B* is introduced to someone with an autoimmune disease, it worsens his or her condition due to its role in inducing a Th1 response (Lee and Buchman 2009; [www.wikipedia.org](http://www.wikipedia.org)). A recent study has found that the *IL12B* genotype is associated with a need for surgery in Crohn’s disease patients – in other words, it is most often found in severe cases of the disease (Dubinsky et al. 2011).

### **IL23R**

The *IL23R* gene, located on chromosome 1, is a cytokine protein involved in the differentiation and maintenance of Th17 lymphocytes (Barrett and Chandra 2011; Hubbard and Cadwell 2011; Lee and Buchman 2009). Individuals with inflammatory bowel disease have high levels of *IL-17*, *IL-21*, *IL-22*, and *IL-23* in their intestines (Hubbard and Cadwell 2011). It has been suggested that mutations in the *IL23R* are protective, making an individual less susceptible to Crohn’s disease (Barrett and Chandra 2011; Clark et al. 2011; Dessein et al. 2008; Newman et al. 2009). The mutation of this gene associated with Crohn’s disease encodes the amino acid change Arg381Gln. This mutation is found in approximately 14% of individuals with European ancestry and confers nearly a 3-fold reduced risk for Crohn’s disease (Cho and Brant 2011).

### **IRGM**

The *IRGM* gene, located on chromosome 5, is involved in autophagy (Barrett and Chandra 2011). This gene, which stands for immunity-related p47 guanosine triphosphatases, confers a predisposition to Crohn’s disease on an individual (Dessein et al. 2008). The *IRGM* gene is related to innate immunity (Lee and Buchman 2009)

## **MIR21**

The *MIR21* gene is a microRNA gene that affects the transcription of many genes. It is an anti-apoptotic factor that is highly expressed in many types of cancers. It is differentially methylated in individuals with Crohn's disease in comparison to controls (Adams et al. 2014a).

## **MSH2**

It has been found that there is increased expression of the *MutS* homolog 2 (*MSH2*) in the mucosal and submucosal tissues of individuals with Crohn's disease, when compared to controls (Floer et al. 2008). This gene, located on chromosome 2, is most often associated with colorectal cancer (Lawes et al. 2005). This is interesting because individuals with Crohn's disease are at a high risk of developing colorectal cancer later in life.

## **NLRP3**

The *NLRP3* gene (NOD-like receptor family, pyrin containing domain 3) is located on chromosome 1q44 and is associated with various autoimmune diseases (Lewis et al. 2011). The protein product of this gene has been shown to interact with the *NOD2* protein (Lewis et al. 2011). Some studies have found an association between the *NLRP3* gene and Crohn's disease; however, a recent study has found there to be no association between the gene and Crohn's disease indicating the possibility of the original study showing false positive results (Lewis et al. 2011). Due to the fact that this refutation is based solely on one study conducted only on United Kingdom patients, additional studies will be necessary to determine whether or not there is an association between the *NLRP3* gene and Crohn's disease in different populations.



### **NKX2.3**

One gene found to be associated with Crohn's disease through association studies is the *NKX2.3* gene located on chromosome 10. It has been found that one haplotype of the *NKX2.3* gene (those individuals with the SNP rs10883365) confers susceptibility to both Crohn's disease and ulcerative colitis by increasing the expression of the *NKX2.3*-mRNA in the colonic mucosa (Arai et al. 2011).

### **PTPN2**

The *PTPN2* (protein tyrosine phosphatase, non-receptor type 2) gene is a signaling molecule gene located on chromosome 18p11 (Festen et al. 2011). It is activated by *IFN-γ* and limits defects in the epithelial barrier (Scharl et al. 2009). When an individual has a mutation in this gene, however, defects in the epithelial barrier occur. Crohn's disease is often associated with increased permeability of the intestinal epithelium, such as is seen in mutations of the *PTPN2* gene (Scharl et al. 2009). This gene (in addition to *IL18RAP*, *TAGAP*, and *PUS10*) is associated with increased risk to both Crohn's disease and celiac disease (Festen et al. 2011). The gene is involved in regulating the immune system and inhibiting T-cell proliferation. *PTPN2* has been shown to limit pro-inflammatory effects in monocytes and epithelial cells of the intestines (Scharl et al. 2011b). Abnormalities of this gene have been shown to enhance T-cell proliferation and increase inflammation, both of which are seen in individuals with Crohn's disease (Marcil et al. 2011; Scharl et al. 2009). A study of pediatric Crohn's disease patients in Canada found a strong association between Crohn's disease and three different SNPs on the

*PTPN2* gene (Marcil et al. 2011). There is an interaction between the *PTPN2* gene and the *NOD2* gene. *PTPN2* is regulated by muramyl-dipeptide (MDP) through a *NOD2*-dependant mechanism and the gene also controls *NOD2*-mediated cytokine secretion and autophagy (Scharl et al. 2011b). The *PTPN2* gene also regulates autophagy and mutations of this gene could increase susceptibility to bacterial invasion, leading to inflammation in the intestines (Scharl et al. 2011a). There is a variant of the *PTPN2* gene that protects against Crohn's disease, but at the same time this variant is a risk factor for developing type 1 diabetes and rheumatoid arthritis (Ventham et al. 2013).

#### **TPMT\*16 and TPMT\*19**

There are two rare allelic variants of the thiopurine *S*-methyltransferase gene: *TPMT\*16* and *TPMT\*19* (Hamdan-Khalil et al. 2005). As the name implies, this gene catalyzes the *S*-methylation of thiopurine drugs. These drugs are typically used in the treatment of leukemia, as well as the treatment of inflammatory bowel diseases (Hamdan-Khalil et al. 2005). Individuals with the *TPMT\*16* and *TPMT\*19* variants had normal *TPMT* activity after taking the thiopurine medications, rather than the decreased enzymatic activity that was expected (Hamdan-Khalil et al. 2005). This elucidates the need for genotyping of individuals prior to the prescribing of medications, as not all individuals will respond the same way to medicine.

#### **OTHER GENETIC FACTORS RELATING TO CROHN'S DISEASE**

While it has been shown on the preceding pages that there are many individual genes that are associated with Crohn's disease, it should also be pointed out that the genetic interaction of many of these genes also influences disease susceptibility. A recent study looked at the

interaction of three SNPs from each of the following genes: *CARD15*, *IL23R*, and *TLR9* for a total of nine SNPs. They found that there was no interaction between *CARD15* and either *IL23R* or *TLR9*, but that there was a significant genetic interaction between the *IL23R* and *TLR9* SNPs (Menoza et al. 2011). It would be interesting to look at all of the genes associated with Crohn's disease and see if there are other genetic interactions out there. If so, this might alter the low heritability of the disease that has been found thus far.

In addition to genes that either protects an individual from developing Crohn's disease or that increases an individual's susceptibility to the disease, there are genes that affect the behavior of the disease. It has been found that different genes may be significantly associated with young Crohn's disease patients than with those who have had the disease for years (Arijs et al. 2011). The genes that were significantly associated with early Crohn's disease were either involved in innate immunity (*C2*, *CFI*, *DUOX2*, and *LCN2*) or related to epithelial barrier function (*MUC1* and *MUC4*), while those genes significant in late Crohn's disease were involved primarily in the immune/inflammatory response (Arijs et al. 2011). SNPs in one gene have been found to be associated with an increased risk of needing surgery (*SERPINA1*) while another (*NOX4*) is associated with a decreased risk of surgical intervention (Ryan et al. 2011). There are also SNPs (in the *CCL2* and *CCL7* genes) that have been found to be associated with recurrence of Crohn's disease after an initial surgical intervention (Siegel et al. 2011).

One study recently used blood samples to test for gene expression in individuals with Crohn's disease, and found that many genes (with various functions) are differentially regulated between individuals with active Crohn's disease and individuals whose Crohn's disease is currently inactive (Burakoff et al. 2010). The following immune response/inflammation genes were found to be up-regulated in individuals with active Crohn's disease: *CD79B*, *C1R*, *GTPBP1*,

*IL18RAP*, and *MRC2*. The following cell adhesion genes were found to be up-regulated in individuals with active Crohn's disease: *MFGE8* and *PKP3*. The following immune response/inflammation genes were found to be down-regulated in individuals with active Crohn's disease: *AIF1*, *ALOX 15*, *ATP6VOA2*, *CTSC*, *CLC*, *IL5RA*, *LGALS12*, *LMAN2L*, *LYZ*, *MGST2*, and *TLR7*. The following cell adhesion genes were found to be down-regulated in individuals with active Crohn's disease: *CTNNAL1*, *CD33*, *CCR3*, *ENG*, *FCGBP*, *ITGA4*, *PTPRS*, *PRG4*, and *TGFB1*. The following cytokinesis genes were found to be down-regulated in individuals with active Crohn's disease: *CETN2*, *C10orf9*, and *NEK3*. It should be noted that all genes listed in this paragraph as differentially expressed were statistically significant at the 0.05 level (Burakoff et al. 2010). As can be seen, some genes with one purpose are up-regulated while others with the same purpose are down-regulated in individuals with active Crohn's disease. It is due to this differential expression that all genes relating to the disease eventually need to be studied in order to better understand the epidemiology of the disease.

The three genes chosen for study in this project are the *AT16L1* gene, the *NOD2* gene, and the *PTPN2* gene. Table 5 shows location of the SNPs and the risk alleles associated with these three genes (Barrett et al. 2008).

*Table 5 The SNP, location, risk allele, and odds ratio associated with the SNPs previously found to be associated with the three genes used in this project (Barrett et al. 2008).*

<b>SNP</b>	<b>Chromosome</b>	<b>Risk Allele</b>	<b>Odds Ratio</b>
<i>ATG16L1</i> - rs3828309	2q37	G	1.28
<i>NOD2</i> - rs2066847	16q12	C	3.99
<i>PTPN2</i> - rs2542151	18p11	G	1.35

As of 2011, there were seventy-one genes associated with disease penetrance of Crohn's disease (Barrett and Chandra 2011; Hubbard and Cadwell 2011). Many of the genes that are associated with increased risk of Crohn's disease are genes that alter proteins to decrease

bacterial resistance, such as *NOD2*, *ATG16L1*, *IBD5*, and *IL23R* (Barrett and Chandra 2011). It should be mentioned that of the seventy-one loci associated with Crohn's disease, each of these variants only confer a minimal degree or risk of disease onset (Hubbard and Cadwell 2011). All 72 susceptibility loci combined only accounts for 33% of the disease risk (Hubbard and Cadwell 2011).

With the many different susceptibility genes associated with Crohn's disease, it has been difficult for scientists to develop a genetic testing method to diagnosis the disease in a patient. One recent attempt used 46 SNPs identified from genome wide association studies to diagnosis 875 Crohn's disease patients (Cleynen et al. 2010). This study used the *NOD2*, *IL23r*, and *ATG16L1* genes, as well as 30 other genes with odds ratios that range from 1.08 to 3.99. They found that individuals with the disease can be clustered into six separate groups, while healthy individuals cluster into five groups that are significantly different than the groups of diseased individuals (Cleynen et al. 2010). After the cluster analysis, they performed a principal components analysis that determined the 6 clusters were best explained by using 3 SNPS on genes *TNFSF15*, *IRGM*, and location 6q23.3. This method of testing was 95.5% accurate in correctly classifying individuals with the disease. It should be noted, however, that 58 individuals with Crohn's disease had to be dropped from the study because they had missing genotypes for at least one of the SNPs used for determination in the study. For healthy controls, 21.5% were incorrectly diagnosed as having Crohn's disease, and an additional 10.6% were unable to be classified due to missing genotype data (Cleynen et al. 2010). The results from this study elucidate the fact that to date there is no genetic study that works well to classify individuals with Crohn's disease without also misdiagnosing individuals who do not have the disease.

Of the multitude of genes associated with the disease, this study will focus on three immune response genes whose alleles are highly correlated with Crohn's disease (*NOD2*, *ATG16L1*, and *PTPN2*). *NOD2* (nucleotide-binding oligomerization domain-containing protein 2) is an intracellular pattern recognition receptor that recognizes bacterial molecules and stimulates an immune reaction in response. Mutations in this gene can result in a decrease in intracellular bacterial sensing (Barrett and Chandra 2011). The *NOD2* gene is activated by viral ssRNA, which then triggers phosphorylation of transcription factor interferon response factor-3 (IRF3). Translocation of the IRF3 leads to the activation of interferon genes, such as IFN- $\beta$ , which mediates innate immune responses to viral agents (Barrett and Chandra 2011). In other words, to successfully fight off a viral infection, the *NOD2* gene must be functioning properly as it is the first step in the pathway that results in an antiviral response. In response to bacterial products, *NOD2* is activated to in turn activate NF- $\kappa$ B, which is a regulator of inflammation (Barrett and Chandra 2011). This activation of *NOD2* is controlled by the leucine rich repeat domain, and mutations in this region result in the inability of cells to be able to recognize muramyl dipeptide, a fragment of peptidoglycan found in the cell walls of both gram-positive and gram-negative bacteria (Barrett and Chandra 2011). If the muramyl dipeptide is not recognized by the leucine rich repeat domain of the *NOD2* gene, the body will not make an immune response to the presence of invading bacteria.

*ATG16L1* (autophagy-related protein 16-like 1) is involved in apoptosis, as well as in helping the immune system destroy some bacteria and viruses. The *ATG16L1* gene encodes a 44-kD protein on chromosome 2 that is involved in the intracellular autophagy complex (Fabio et al. 2011b). Mutations in this gene is thought to diminish bacterial clearance by impairing the autophagy ability of the cell (Barrett and Chandra 2011; Dessein et al. 2008). In order to

understand what this means, it is important to first get a good definition of what autophagy is. “Autophagy degrades damaged organelles and proteins, in homeostasis and as a response to starvation, and is important for the clearance of pathogens (xenophagy), which is required for immunity to multiple different types of bacteria” (Cho and Brant 2011). Autophagy is necessary in the mediation of the inflammatory response. It works by capturing organelles and cytoplasm and holding them within a membrane-bound organelle known as an autophagosome; once held, a lysosome comes along and destroys the microorganism within the autophagosome through the process xenophagy (Barrett and Chandra 2011). If this process is disrupted, the result is a production of inflammatory cytokines, which may lead to the poor response to intracellular bacteria that is found in individuals with Crohn’s disease (Barrett and Chandra 2011). It has been found that several genetic variants associated with an increased risk of Crohn’s disease are associated with autophagy (Hubbard and Cadwell 2011). The T300A mutation in the *ATG16L1* gene, which is associated with Crohn’s disease, yet confers a protective effect in ulcerative colitis (Serbati et al. 2014). One study found that the *ATG16L1* risk variant was associated with an increase in *E. coli* proliferation in the body (Elliott et al. 2011). A mutation in the *ATG16L1* gene could also result in the invading pathogens (such as *Salmonella*) not being held within the autophagosome, and thus not destroyed by the lysosome. Specifically, “*ATG16L1* mutations cause disordered production of key anti-microbial peptides made in Paneth cells in the lining of the intestine” (Barrett and Chandra 2011). The odds ratio of this gene in association with Crohn’s disease depends on the number of copies of the Thr300Ala A→G variant the individual has. The odds ratio for the Thr/Thr genotype is 0.76 in a Mediterranean population; for Thr/Ala genotype, the odds ratio is 0.89; for the Ala/Ala genotype, the odds ratio is 1.42 that the individual will develop Crohn’s disease (Fabio et al. 2011b). In a European population, it was found that having

one copy of the T300A(+) variant gives an odds ratio of 1.26 (OR of 1.00 for T300A(-) variant) (Fabio et al. 2011b). These results show that the odds ratio of this gene (as well as most other genes) is entirely dependent on the population under study.

*PTPN2* (protein tyrosine phosphatase, non-receptor type 2) regulates epithelial barrier function, particularly in the intestines. Less is known about the *PTPN2* (protein tyrosine phosphatase, non-receptor type 2) gene than about the other two genes being used for this study. What is known is that *PTPN2* (protein tyrosine phosphatase, non-receptor type 2) regulates epithelial barrier function, particularly in the intestines. The *PTPN2* gene is a signaling molecule gene located on chromosome 18p11 (Festen et al. 2011). This gene (in addition to *IL18RAP*, *TAGAP*, and *PUS10*) is associated with an increase risk to both Crohn's disease and celiac disease (Festen et al. 2011).

If it is found that there are significantly different methylation levels in individuals with Crohn's disease, using buccal swabs to test for methylation levels may be able to be used as a non-invasive diagnostic test. To date, there are no reliable biomarkers that are specific for Crohn's disease (Karatzas et al. 2014). By studying the methylation of these genes, it is hoped that Crohn's disease can be more fully understood, and hopefully more effective diagnostic tools and treatments can eventually be devised. Other environmental factors (smoking, toxins, alcohol consumption, ethnicity, pollution, stress, other diseases, etc....) will also be examined. The proposed research will help to elucidate the role of epigenetics on the immune system. This research takes a holistic view to the immune system that is not often seen in most epigenetic studies. Rather than looking solely at whether or not there is an epigenetic effect on the immune system, this research will also explore the evolution of an autoimmune disorder, the selection



forces acting upon the genes, the environmental factors that can affect the expression of the genes, and the functionality of the genes subsequent to epigenetic factors.

The genetic mutations that are associated with Crohn's disease are found primarily in industrial societies (Barrett and Chandra ; Hermon-Taylor ; Nakazawa). This fact leads to some interesting anthropological questions. What is it about such societies that would lead to these mutations? What environmental risk factors are found in industrial societies that are not found in other societies? How has living in a modern society and being exposed to a different diet, different stresses, and increased antibiotic usage led to an increased prevalence of Crohn's disease? Is natural selection operating on the genes in question? If so, what is causing this selection?

These are the questions that will be addressed through this proposed research. One of the ways in which this will be done is by looking at the epigenetic methylation profiles of individuals with Crohn's disease. As a person ages, the amount of DNA methylation found in the genome increases. Monozygotic twins have been found to have methylation profiles that diverge as they age (Backdahl 2010; Barres and Zierath 2011). These different methylation profiles may help to explain the discordance of Crohn's disease phenotype that has been found in multiple twin studies.

The twenty-four genes or regions in Table 6 either have SNPs that are believed to be associated with Crohn's disease or are differentially expressed in individuals with active Crohn's Disease. Some of these genes, such as the NOD2 gene, are known to have multiple SNPs that are associated with Crohn's Disease.

*Table 6 List of genes with either SNP or methylation profile associated with Crohn's disease.*

GENE	FUNCTION	REFERENCE
NOD2	Capsase recruitment protein; immune response	(Hugot et al. 2001; Sventoraityte et al. 2010)

IL23R	Cytokine receptor	(Dessein et al. 2008; Glas et al. 2007; Lee and Buchman 2009; Newman et al. 2009; Sventoraityte et al. 2010)
ATG16L1	Autophagy protein	(Dessein et al. 2008; Lee and Buchman 2009; Newman et al. 2009; Sventoraityte et al. 2010)
IRGM	Autophagy pseudogene	(Dessein et al. 2008; Lee and Buchman 2009; Parkes et al. 2007)
MST1	Hepatocyte growth factor-like protein	(Lee and Buchman 2009)
NKX2-3	Homeobox gene	(Franke et al. 2008; Parkes et al. 2007)
MSH2	<i>E coli</i> mismatch repair gene	(Floer et al. 2008)
TLR5	Cytokine receptor; immune response	(Stanislawowski et al. 2009)
PTPN2	Signaling enzyme; protects epithelial barrier function	(Scharl et al. 2009)
BPI	Bacteria permeability increasing protein	(Klein et al. 2005)
TNF- $\alpha$	Cytokine involved in inflammation	(Snoeks et al. 2009)
CARD15	Capsase recruitment protein; immune response	(Glas et al. 2007; Lee and Buchman 2009)
IL12B	Cytokine; immune response	(Lee and Buchman 2009)
STAT3	Transcription activator in response to cytokines	(Lee and Buchman 2009; Scharl et al. 2009)
TPMT	Metabolism enzyme	(Hamdan-Khalil et al. 2005)
5p13region	Site of many immune system genes, including IL7 receptor and prolactin receptor	(Libioulle et al. 2007; Parkes et al. 2007)
10q21 region	Site of tumor suppressing genes	(Franke et al. 2008)
FOXO3	Tumor suppressor	(Snoeks et al. 2009)
ST13*	Tumor suppressor	(Burakoff et al. 2010)
ITGA4*	Immune response	(Burakoff et al. 2010)
ATP6V0A2*	ATPase enzyme	(Burakoff et al. 2010)
MYB*	Proto-oncogene protein	(Burakoff et al. 2010)
AIF1*	Inflammation; immune response	(Burakoff et al. 2010)
IL18RAP*	Interleukin receptor; immune response	(Burakoff et al. 2010)

## *HUMAN EPIGENETICS*

How long is the average human lifespan? Is it coded for in our genes, or is it the result of our environment? In general, people live longer today than they did 100 years ago, and those individuals lived longer than people lived 500 years ago. Many people are under the misconception that our bodies are evolving to live longer and longer. This is not the case. In terms of evolution, our genome does not care whether or not humans have long life spans. Natural selection does not act on the genome in order to favor a long lifespan, but rather to increase reproductive fitness. Therefore, those genes most beneficial to making sure we survive to reproductive age are selected for, while others (including whether or not we live to be 100) may be selected against. One gene may be beneficial in terms of surviving to reproductive age, yet that same gene may shorten the overall lifespan of the individual. Evolutionary change in a species is a long process. Our increasing lifespan cannot be explained by genetic changes that have taken place in only a few hundred years. In terms of evolutionary change, that short of a timescale is not sufficient to have evolved genes within our species promoting a longer lifespan. The reason humans are living longer is not in our genome, but rather in our environment. With more effective medicines and better nutrition, people are able to live much longer. While our environment (in the forms of nutrition, medicine, and lifestyle) does not always affect our genome, it does have a significant effect on our epigenome. This paper will explore the epigenome, the factors that affect it, and the disease consequences that can occur to humans due to epigenetics.

## *HISTORY OF EPIGENETICS*

Epigenetics is a fairly recent field of study that is sometimes referred to as soft inheritance. The field of epigenetics has thus far mostly been studied on plants and insects, but

this paper will focus primarily on human epigenetics. “Since the first evidence of transgenerational epigenetic inheritance in plants, the idea has been put forward that epigenetic marks, such as DNA methylation, histone modifications and small RNAs, could contribute to transgenerational inheritance of acquired traits in metazoans, as well” (Brasnet and Chambeyron 2013). There are many different definitions to explain epigenetics. The first definition of epigenetics was in 1942, and it states that epigenetics is “the casual interactions between genes and their products that allow for phenotypic expression” (Waddington 1942). This definition has evolved and expanded over time. One author states that epigenetics is “a heritable state of gene expression that is not due to changes in the DNA sequence” (Barres and Zierath 2011). According to dictionary.com, epigenetics is “the study of heritable changes that occur without a change in the DNA sequence”. While these definitions are correct, they are lacking in substance. A more complete definition was given by Heidi Ledford and says, “epigenetics is mostly the study of heritable changes that are not caused by changes in the DNA sequence; to a lesser extent, epigenetics also describes the study of stable, long-term alterations in the transcriptional potential of a cell that are not necessarily heritable” (Ledford 2008). What all this essentially means is that research in the field of epigenetics examines gene expression that may or may not be heritable, but that cannot be explained solely by the genetic make-up of the individual. Environmental changes, stress, and diet may cause metabolic and cellular disorders in mammals that can be transmitted through multiple generations (Brasnet and Chambeyron 2013). The challenge in studying epigenetics is to determine which traits are epigenetic, rather than a consequence of heritable genetic changes. The field of epigenetics includes the interaction of the genome, environmental factors, the aging process, and disease pathogenesis (Itin and Burger 2009).

Researchers in epigenetics study gene expression that may or may not be heritable, but that cannot be explained simply by the genetic make-up of the individual. This includes the interaction of the genome, environmental factors, the aging process, and disease pathogenesis (Itin and Burger 2009). Many human diseases are multifactorial in origin, which means they are the result of interactions between the genetics of the individual and the environment in which he/she lives. The environment can directly affect the epigenetic profile of an individual, which in turn can affect the expression of that person's genetics. Thus, epigenetics may be the ultimate link between genetics and the environmental components that have been shown to play a role in many diseases, particularly autoimmune diseases (Backdahl 2010; Itin and Burger 2009; Lin et al. 2010). Epigenetics may also explain differential susceptibility to diseases in different populations (Andraos et al. 2011). If a specific gene in one population is over-expressed or under-expressed due to an epigenetic cause, then the individuals in that population may be more or less likely to have a disease that is associated with the expression of that gene. Diseases with late-onset phenotypes, such as autoimmune diseases, “involve interactions between the epigenome, the genome, and the environment” (Feinberg 2007). There are two classes of epigenetic diseases – those diseases involving genes that are epigenetically regulated (such as imprinted genes) and those diseases that are caused by epigenetic changes that affect the entire epigenome (such as DNA methylation). See Figure 8 for a diagram of the interaction of genetics, epigenetics, and the environment.

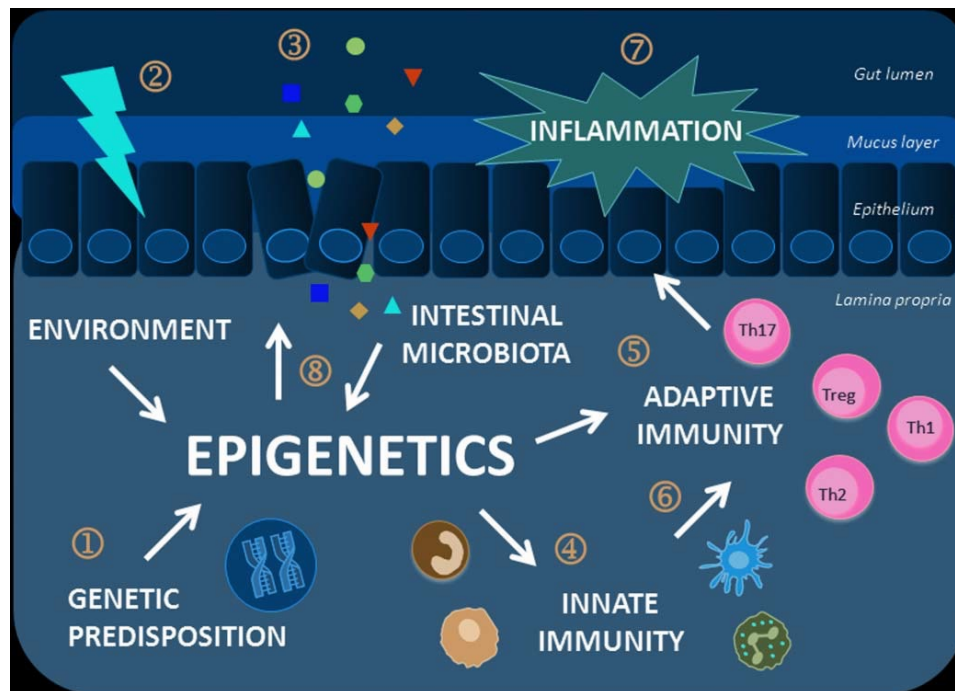


Figure 8 “Roles for epigenetics in IBD pathogenesis. Epigenetics may interact with genetic factors (1), environmental factors like nutrition, smoking, stress or hygiene, and the intestinal microbiome (3) in affecting the immune system. Phagocytic cells within the lamina propria (e.g., monocytes/macrophages, DCs, neutrophils) and epithelial cell barrier represent the central components of the intestinal innate immune system (4). Antigens reaching the lamina propria activate innate immune cells followed by a response of T lymphocytes (Th, T-helper cells; Treg, T-regulatory cells), which represent the key cell population of the adaptive immune system (5). The innate and adaptive immunity arms are tightly cross-regulated serving to uphold intestinal homeostasis and thus to control the complex commensal– host crosstalk (6). The subsequent immune response has consequences on the initiation, resolution, and/or progression of intestinal inflammation (7). Bacteria can also regulate the epithelial gene expression through epigenetic mechanisms (8)” (Dabritz and Menheniott 2014).

Jean Baptiste Lamarck gave one of the first proposed hypotheses for evolution in 1801. He hypothesized a “use it or lose it” view of evolution. He believed that increased use of a part of an organism would cause an increase in the size of that part and that this increase would be inherited by the next generation. Likewise, unused parts of the organism would shrink over time from disuse. These beliefs became known as the transmission of acquired characteristics. In 1859, Charles Darwin published his book On the Origin of Species, which proposed the theory of natural selection to explain evolution. With his work and that of later scientists, the Modern

Evolutionary Synthesis was created. The Modern Evolutionary Synthesis (AKA the Modern Synthesis) states that: heredity occurs through the transmission of germline genes; inherited variation is found as variation in the DNA base sequence; these variations were either pre-existing or the result of new mutations; selection acts on the variation; heritable variations have small effects, resulting in a slow and gradual evolution; and evolution is based on vertical descent (Jablonka and Lamb 2008). This view is now being challenged. As scientists learned more about the field of genetics and natural selection (the Modern Synthesis), the Lamarckian theory of the inheritance of acquired characteristics was believed to be false. With the emerging field of epigenetics, however, there is evidence supporting the theory that traits acquired during one's lifetime can be inherited by his/her offspring through epigenetic mechanisms. This has led to a revived interest in the Lamarckian idea of the inheritance of acquired characteristics (Curley et al. 2011).

Many scientists opposed using a Lamarckian theory when looking at epigenetics. Feinberg and Irizarry proposed a non-Lamarckian theory for the role of epigenetics in evolution. They propose a theory of stochastic variation. They state, “genetic variants that do not change the mean phenotype could change the variability of phenotype; and this could be mediated epigenetically. The data suggests that genetically inherited propensity to phenotypic variability, even with no change in the mean phenotype, substantially increases fitness while increasing the disease susceptibility of a population with a changing environment” (Feinberg and Irizarry 2009). This explanation provides a mechanism to explain the role of epigenetics in developmental biology, as well as to explain previously unexplained variations found in complex diseases. The fundamental tenets of Darwinian inheritance are still true, but epigenetic inheritance “expands

the range of options available to genes”, while at the same time evolutionary adaptation is still the product of natural selection (Haig 2006).

According to Jablonka *et al.*, the challenge to the Modern Synthesis involves five new tenets (Jablonka and Lamb 2008). (1) The inheritance of traits is dependent on more than just our DNA. There are non-DNA variations that can shape evolutionary change and genetic evolution. ([www.gen.tcd.ie/molpopgen/link%20files/Ireland\\_mtDNA.xls](http://www.gen.tcd.ie/molpopgen/link%20files/Ireland_mtDNA.xls)) Soft inheritance (epigenetics) exists and is important to the study of human evolution. This soft inheritance is comprised of the above-mentioned non-DNA variations, as well as developmentally induced variations in DNA sequence that occur throughout a person’s lifetime (acquired characteristics). (3) Epigenetic modifications are possible targets of natural selection. (4) Saltational changes (sudden and drastic changes) beyond the species level that lead to evolution are common. (5) The original Tree of Life pattern of variation fails to explain all of the similarities and differences between species. For the purposes of studying epigenetics, the first two tenets (non-DNA variations can shape evolution and epigenetics is the means by which this occurs) are of great importance.

#### *EPIGENETIC MODIFICATIONS: DNA METHYLATION*

While genes are contained within the genome, the activity of a gene is often contained within the epigenome. The epigenome is comprised of the chromatin structure, non-coding RNAs, and the DNA methylation present on the genome. While the genome is the same throughout the body (with the exception of immune cells experiencing V(D)J recombination), the epigenome varies between different cell types, and possibly even between two cells of the same cell type (Cooney 2007). V(D)J recombination is the mechanism of genetic recombination that occurs in lymphocytes during the early stages of T cell and B cell maturation. Similar to the



genome, the epigenome is duplicated during mitosis and meiosis. Where the genome can only change through mutation, the epigenome can change through stochastic epimutation or directed epimutation, causing a new pattern of gene expression. The mechanisms of epigenetics are affected by: development (in *utero* and during childhood), environmental chemicals, drugs/pharmaceuticals, aging, and diet (Reece 2012). Epigenetic mechanisms have been linked to cancer, autoimmune diseases, mental disorders, and diabetes.

There are many different forms of epigenetic modification than can occur. One form of epigenetic modification is DNA methylation. One of the key epigenetic mechanisms by which cells regulate gene expression is through DNA methylation. “Among its roles are the dynamic regulation of gene expression, for example, as part of an evolving immune response, and cell differentiation in specialized tissues” (Cooke et al. 2012). Of all the different types of epigenetic modification, DNA methylation is one of the most stable forms of modification (Backdahl et al. 2009; Barnett et al. 2010). Due to the fact that methylation is the most common epigenetic modification that is preserved through mitosis, it can be seen to act as an epigenetic memory for the epigenome (Backdahl et al. 2009). Some scientists believe that restoring the original DNA methylation profile of an individual may reverse certain diseases that were caused by epigenetic changes. DNA methylation is the most common epigenetic modifier in the human genome (Backdahl 2010).

DNA methylation occurs by the attachment of a methyl group (CH<sub>3</sub>) to a cytosine at position 5, forming 5-methyl-2'-deoxycytidine (Barnett et al. 2010; Brooks et al. 2010; Gupta et al. 2010). This is an important epigenetic marker in CpG dinucleotides of mammalian cells. CpG dinucleotides are places within the genome where a cytosine is followed by a guanine in the genetic sequence. A sequence of three nucleotides makes up a codon that codes for a particular

amino acid. There are 64 possible codons and these code for 20 different amino acids, meaning that multiple codons can all code for the same amino acid. In a protein, it is usually the arginine amino acid that is methylated (CGU, CGC, CGA, or CGG) (Walsh 2006). There may be either one or two methyl group additions to each arginine amino acid. DNA methylation can also occur at the N-4 positions of cytosine and the N-6 position of adenine, though these are less common. The methyl donor is S-adenosylmethionine, which is a product of methyl metabolism (Cooney 2007). The cannibalizing of folic acid, methionine, betaine, and choline metabolizes methyl groups. Methyl groups are transported and transferred by folate, methionine, zinc, and vitamin B<sub>12</sub> (Cooney 2007).

Most DNA methylation occurs at cytosines at CpG sites, but not at CpG islands. A CpG island is a small stretch of 300-3000 base pairs that contains more than 50% CpG content. These islands are most often found in the promoter region of the gene. Approximately 70% of promoter regions contain CpG islands, as compared to about 1% of the remaining genome (Reece 2012). Hypermethylation of the promoter region is associated with inactivation of that gene (Barnett et al. 2010). In other words, the more methyl groups that are attached to the promoter region of a gene, the less likely that gene is to be transcribed and expressed. The combination of CpG islands primarily being found in promoter regions and inactivation of the gene when the promoter region is hypermethylated is why methylation of cytosines within a CpG island is rarer than in other areas. Methylation is mutagenic away from CpG, thus hypomethylated regions retain CpG islands.

A recent study looked at allele-specific methylation across the genome, particularly in those regions associated with complex diseases. The researchers found a lot of inter-individual variation, but that 5% of the sites they analyzed had allele-specific methylation (Hutchinson et al.

2014). Most of the sites studied were in high linkage-disequilibrium. They also found that 81% of the methylated sites were under genetic influence. They performed this study using peripheral blood mononuclear cells that were put through a microarray for analysis. The results were confirmed using bisulfite pyrosequencing.

While a gene is discrete in that one DNA sequence codes for a specific protein, DNA methylation is quantitative in that a little methylation may have no effect while a lot of methylation may cause either inactivation of the protein or cause the protein to not be synthesized in the first place. This is due to the fact that methylation of cytosine residues impairs the binding of transcription factors, which results in reduced gene expression (Cooke et al. 2012). It is interesting to note that much of the non-coding region of the genome is methylated (Callinan and Feinberg 2006). It is possible that these regions used to code for something in our evolutionary past, but that the accumulated methylation over time caused the genes that used to be found in these regions to become non-functional. This supports the Lamarckian view that a lack of use will cause that part to shrink or become discontinued in subsequent generations. It should be noted, however, that Lamarck was talking about characters, not genes. These methylated non-coding regions may help scientists to better understand the function of the DNA regulatory sequences that are found there. DNA methylation is dynamic and can control the timing of cellular events (Gupta et al. 2010).

The methylation of a gene recruits methyl-binding proteins that then recruit chromatin-remodeling proteins, which can then modify histones and thus form inactive chromatin (Barres and Zierath 2011). The changes in chromatin often results in spatial changes within the cell where the active genes are usually located towards the nucleus of the cell, while inactive genes tend to be closer to the membrane. The location of the nucleus does not change, but rather the

histones slide around and cause chromatin remodeling. The literature suggests that chromatin remodeling is especially pronounced in the elderly, indicating that the number of histone modifications increase throughout a person's lifetime (Brooks et al. 2010). Promoters of transcriptionally active genes are usually hypo-methylated (Backdahl et al. 2009; Burdge and Lillycrop 2010). This makes sense due to the fact that hypermethylation would cause these genes to become inactive. It should be noted that there are exceptions to the hypermethylated/off, hypomethylated/on theory. For example, if a methylated CpG island falls within the response element of a transcriptional repressor, this can cause the gene to become expressed since it turns off the repressor rather than turning off the gene itself (Burdge and Lillycrop 2010). In fact, approximately 70% of CpG dinucleotides found in heterochromatin and retrotransposons are methylated (Burdge and Lillycrop 2010). One may think that hypermethylation of a gene other than a transcriptional repressor is a bad thing since it turns off a gene. This is not always the case. DNA methylation is primarily a means of genome control, such as suppressing the expression of intragenomic parasitic sequences like endogenous retroviruses (Cooney 2007).

A transferase is a class of enzyme that works to transfer a specific functional group from a donor molecule to an acceptor molecule. There are at least five methyl transferases that carry out the addition of methyl groups to DNA. The amount of methylation present at any given CpG site is maintained throughout the cell replication process by DNA methyl transferase 1 (*DNMT1*), which preferentially methylates hemi-methylated substrates (Brooks et al. 2010; Burdge and Lillycrop 2010). There are actually two isoforms of *DNMT1*. *DNMT1o* is found in oocytes and early embryos, where it maintains methylation of imprinted genes, but it is not found in adult tissues. *DNMT1s* is also found in oocytes and early embryos, but it is also present in adult tissues, although at lower levels than that found in oocytes and early embryos (Burdge and Lillycrop

2010). *DNMT2* works to methylate the aspartyl-tRNA during protein biosynthesis. New methylation is controlled by DNA methyl transferase 3a (*DNMT3a*) and DNA methyl transferase 3b (*DNMT3b*) (Backdahl et al. 2009; Brooks et al. 2010). The final DNA methyl transferase involved is *DNMT3L*, which does not have catalytic activity itself but rather binds to *DNMT3a* and *DNMT3b* in order to regulate their functions (Brooks et al. 2010; Burdge and Lillycrop 2010).

It should also be noted that the location of the gene could also affect the epigenome. In twin studies, it has been found that epigenetic modifications (such as DNA methylation) are more likely to occur in gene-poor regions than in gene-rich regions (Choi and Kim 2007). These gene-poor regions of the genome were found to be especially sensitive to environmental triggers. It is possible that the reason DNA methylation is more common in gene-poor regions of the genome is that there is already an excess of methylation found in these non-coding regions of the genome, as was stated earlier.

### *Genetic Imprinting*

Methylation of genes has been shown to lead to genetic imprinting (Anway et al. 2005). Genetic imprinting is when only one of the two inherited copies of a gene (one from each parent) is expressed. This happens when one copy of the gene is silenced due to the addition of methyl groups during egg or sperm development, which occurs approximately halfway through human gestation.

Imprinted genes are usually found near differentially methylated regions – the amount of methylation on the chromosome from one parent is more than the amount inherited from the other parent (Tollefsbol 2010). One example of a disease caused by genetic imprinting is

Beckwith-Wiedemann syndrome, which is characterized by prenatal overgrowth, a midline abdominal wall, cancer, and other possible malformations (Feinberg 2007). Disrupted imprinting on chromosome 11 causes this syndrome. Other examples of genetic imprinting disorders are Prader-Willi syndrome and Angelman syndrome, both of which are associated with mental retardation (Feinberg 2007). Prader-Willi syndrome and Angelman syndrome are caused by a deletion on chromosome 15q11-13. The phenotype differs depending on whether the deletion was transmitted through the child's mother or the child's father (Rando 2012).

Genes that are imprinted are more susceptible than other genes to environmental effects due to the fact that there is only one working copy of the gene, with no "back-up" as is found in most genes. Since imprinting happens during egg or sperm formation, many things that the mother are exposed to (such as diet, hormones, and toxins) can impact the expression of these genes into the next generation and beyond (Jaenisch 1997). Genetic imprinting is currently the only known epigenetic mechanism that influences germline transmission (Anway et al. 2005). Genetic imprinting is one reason why it is so dangerous for pregnant women to be exposed to harmful toxins. What a woman does with her body during the nine months of her pregnancy can have health implications for her descendants for generations to come.

#### *Maternal Effect Genes*

According to biology-online dictionary ([www.biology-online.org](http://www.biology-online.org)), a maternal effect gene is "a gene from the mother's genome in which its phenotype in the zygote is influenced from the mother's genotype, not the zygote's". A maternal effect occurs when the offspring exhibits the phenotype of the mother, even though the offspring's genotype may be different. Essentially, a maternal effect gene is one where a genetic mutation on a gene in the mother influences the development of the offspring, even though the genetic mutation itself is not passed on to the

offspring. Each maternal effect gene may have a pleiotropic effect on the offspring. This maternal effect can occur due to mRNA or proteins that the mother supplies to the egg *in utero* prior to the initiation of zygotic gene expression or may be caused by factors in the maternal environment. During oogenesis, thousands of maternal mRNA transcripts are put into the egg (Wade 2001). Many different species have been shown to have traits that are influenced by the maternal genome rather than by the offspring's genome (Varona et al. 2015).

There is recent hypothesis called the developmental origins of health and disease (DOHaD) hypothesis that states that the intrauterine environment of the offspring may induce changes in organ functions of the offspring that may predispose one to disease in adulthood (Nielsen et al. 2016). In other words, maternal environmental factors during early embryonic development may lead to chronic diseases later in life. The intrauterine environment may be affected by the maternal genome, as well as by the environmental factors affecting the mother, such as smoking, medications taken by the mother, caloric restricted diet, etc.... DNA methylation has been found to have a key function in the DOHaD hypothesis. According to this hypothesis, the intrauterine environment may provide either negative or positive outcomes for the offspring later in life, as well as for future generations. In some cases, the mother's genome and environment may cause the offspring to develop diseases such as diabetes mellitus later in life, while in other cases the intrauterine environment may provide the offspring with an adaptive advantage, such as the ability to survive better in times of famine (Nielsen et al. 2016).

The maternal environment can have a huge impact on the phenotype of the offspring. The effects of the maternal environment on the offspring can be seen in changes in gene expression, which is often influenced by DNA methylation (Champagne and Curley 2009). DNA methylation is inherited through mitosis from mother to daughter cells. DNA methylation is a

stable epigenetic marker with a 0.1-3% methylation loss in daughter cells per mitosis (Champagne and Curley 2009). Due to the stability of DNA methylation and the small loss of methylation in future cell divisions, epigenetic markers can be transmitted through many generations (Panchenko et al. 2016).

In addition to maternal effect genes affecting the phenotype of the offspring, they also may play an important role in speciation (Wade 2001). This can occur when there are negative interactions between the expressed genes from the mother and the expressed genes of the zygote. It has been shown that maternal effect genes are responsible for reproductive isolation between different species of nematodes and mice (Wade 2001). Thus, “maternal effect genes can be viewed as heritable environments” (Wade 2001). The maternal environment can affect the direction, rate, and duration of adaptive evolution (Thiede 1998).

While genes are contained within the genome, the activity of a gene is often contained within the epigenome. The epigenome is comprised of the chromatin structure, non-coding RNAs, and the DNA methylation present on the genome. Maternal effect gene Y causes differential methylation of gene Z in the offspring. Maternal effect gene Y is a genetic condition in the mother that is not present in the offspring. It does, however, have an epigenetic effect on the offspring – it causes differential methylation of gene Z. This differential methylation of gene Z is now part of the epigenome of the offspring, and can be inherited by their future offspring. Thus, the original basis of differential methylation in gene Z is genetic from the mother’s genome. It then becomes epigenetic in the offspring, as it is the epigenome that is being inherited by future generations, rather than the original genetic basis of gene Y.

#### *X Chromosome Inactivation*



Another form of epigenetic control is observed with X chromosome inactivation. During embryonic development of females, one of the two X chromosomes becomes inactivated. The inactive chromosome is made primarily of heterochromatin and is known as the “Barr body” (Tollefsbol 2010). In most females, approximately 50% of the active X chromosomes came from the mother and approximately 50% came from the father (Brix et al. 2009). However, this is not always the case. Some females have skewed X chromosome inactivation where more cells with the X chromosome come from one parent than the other (60-90% rather than 50%). This can be dangerous in cases of X-linked diseases.

Many studies have found that patients with autoimmune thyroid disease have a higher frequency of skewed X chromosome inactivation than is found in controls (Brix et al. 2009). What this means is that these individuals had 60-90% of their X chromosomes be carriers of the disease, giving a much greater chance of the normal chromosomes being inactivated. Thus, they were more likely to express the disease than individuals who had a balanced X inactivation. This might partially explain the predisposition of women to some autoimmune diseases (Brooks et al. 2010)

#### *Telomere Methylation*

The ends of linear chromosomes are protected from genomic instability by telomeres. Telomeres are DNA-protein complexes made up of repetitive TTAGGG sequences. They are up to 20 kilobases (Connor and Akbarian) in length and have a G-rich single-stranded overhang on their 3' end (Koziel et al. 2011). This overhang connects with the repetitive sequences to form a t-loop that protects the telomere. Even though the telomere is protected, it shortens over time as it is continually replicated throughout a person's lifetime. The older an individual is, the shorter telomeres they have.

Telomerase is a reverse transcriptase that helps to extend and maintain the telomeres. Expression of telomerase is usually only found in high amounts in embryonic cells, adult male germline cells, stem cells, and cancer cells. In other words, it is primarily found where the cells are undergoing increased replication (Koziel et al. 2011). The promoter region of telomerase is rich in CpG sites, making it a target for methylation.

DNA methylation regulates telomerase activity. The area proximal to the promoter of telomerase is heavily methylated, and it has been found that increased methylation of this region allows for increased activity of telomerase (Koziel et al. 2011). Telomerase that is functioning incorrectly (either through DNA methylation, genetic mutations, or some other cause) leads to a number of diseases such as: dyskeratosis congenital, aplastic anemia, idiopathic pulmonary fibrosis, and acute myeloid leukemia (Koziel et al. 2011).

#### *EPIGENETIC MODIFICATIONS: HISTONE ACETYLATION*

Histone acetylation also plays a role in the expression of a gene (Backdahl et al. 2009). Histones are the proteins that package and order the DNA into nucleosomes, which are the structural units of chromatin. Histones are the primary protein found in chromatin, around which DNA winds. Whereas the addition of a methyl group to a gene causes loss of gene expression, the opposite is true in the addition of an acetyl group to a histone. Histone acetylation activates transcription and DNA repair. Histone deacetylases remove acetyl groups, and this is usually associated with loss of gene expression or gene silencing (Backdahl et al. 2009; Gavin et al. 2009). This represses the transcription of a gene by restricting access to the promoter regions. Histone acetylations are less stable than CpG methylation, and they require histone acetyltransferase (*HAT*) and histone deacetylase (*HDAC*) to maintain their status. Histones can act as carriers of epigenetic information (Ragunathan et al. 2014).

Histone acetylation is not the only epigenetic event that can happen to a histone. There are other types of histone modifications (methylation, phosphorylation, ubiquitination, and sumoylation), all of which can affect the function of the histones (Callinan and Feinberg 2006). Methylated histones are associated with activation, elongation, or repression of gene expression depending on which amino acid in the sequence the methylation occurs. Histone phosphorylation activates transcription of genes. Histone ubiquitination can silence gene expression and either inhibit or activate transcription of methylated genes. Histone sumoylation opposes acetylation and ubiquitination and can repress gene transcription. See Table 7 for summary of histone modifications. A recent study has proven that histone post-translational modifications can be transmitted independently of specific DNA sequence, DNA methylation, or RNA sequence (Ragunathan et al. 2014).

*Table 7 List of histone modifications and their effects on gene expression, transcription, and DNA repair.*

<b>MODIFICATION</b>	<b>GENE EXPRESSION</b>	<b>TRANSCRIPTION</b>	<b>DNA REPAIR</b>
Acetylation	Expressed	Activates	Activates
Methylation	Expressed/Silenced	Activates/Inhibits	Activates
Phosphorylation	Expressed	Activates	Required
Ubiquitination	Silenced	Activates/Inhibits	No known role
Sumoylation	Silenced	Inhibits	No known role

There is a link between DNA methylation and histone modifications. This link is found in the form of four proteins. Methyl CpG-binding protein 2 (*MeCP2*) and methyl CpG-binding domain (*MBD*) proteins 1, 2, and 3 work to bind methylated DNA sequences and then recruit histone-modifying enzymes (Burdge and Lillycrop 2010). These recruited enzymes work to repress heterochromatin and silence genes. When a cytosine is methylated in a promoter region and bound to *MeCP2*, this can induce dimethylation and trimethylation on histone H3, which then converts euchromatin to heterochromatin. There is also a reciprocal relationship between

DNA methylation and histone modification in that *DNMT1* can be recruited by histone-modifying enzymes (Burdge and Lillycrop 2010).

#### *RNAs AND EPIGENETICS*

Ribonucleic Acid (RNA) is similar to DNA in that both are assembled as a chain of nucleotides. RNA, however, is single-stranded and contains the base uracil instead of thymine. RNA molecules can catalyze biological reactions and control gene expression. Small RNAs are key regulators of chromatin structure (Holoch and Moazed 2015). They have roles in RNA degradation and translational expression. They can modify chromatin. Small RNAs can also target gene expression through RNA interference pathways. Nuclear RNA interference (*RNAi*) pathways have been shown to mediate histone and DNA methylation events that repress transcription in many instances (Holoch and Moazed 2015). The *RNAi* pathways were first discovered while studying a species of yeast called *Schizosaccharomyces pombe*. It was found that deletion of any of three genes (*ago1+*, *dcr1+*, and *rdp1+*) in *S. pombe* resulted in heterochromatic gene silencing and a reduction of the levels of histone methylation (Volpe et al. 2002). This suggested that *RNAi* has an important role in the initiation of heterochromatin formation, as well as the subsequent maintenance of it (Holoch and Moazed 2015).

Long non-coding RNAs are transcripts more than 200 nucleotides in length that do not code for a protein. Long non-coding RNAs are involved in down-regulating gene expression, as well as regulating histone modifications (Spitale et al. 2011). In one famous study, it has been found that “induction of RNA interference in *C. elegans* results in heritable RNA-mediated gene silencing for approximately four to five generations” (Rando 2012). Studies have shown that histone modifications can be affected by RNAs. Small RNAs and long non-coding RNAs can direct both histone modifications and cytosine methylation (Rando 2012). Modification of

histones or of the DNA itself through short non-coding RNAs and other factors can make a gene more or less accessible to transcription. It has been shown in potatoes that small RNAs can cause transmissible transcriptional gene silencing that is associated with promoter methylation (Brasnet and Chambeyron 2013). Long non-coding RNAs contain signals that recruit chromatin-remodeling complexes, such as the X inactive specific transcript (*XIST*) that coats the entire inactive X chromosome in females (Holoach and Moazed 2015). Thus, RNAs are responsible for X chromosome inactivation in female mammals. *XIST* functions by directing the Polycomb repressive complex 2 (PRC2) H3K27 methyltransferase to chromatin. The long non-coding RNAs are believed to act as scaffolds for Polycomb recruitment in a complex system of assembled components, most likely involving the contributions of other protein factors and histone modifications, as well (Holoach and Moazed 2015).

The maintenance of epigenetic information is thought to be ensured through self-reinforcing positive feedback loops (Holoach and Moazed 2015). These loops are formed by the functional coupling of different signaling events, such as when methylation events are physically coupled to proteins that recruit small interfering RNA (siRNA) amplification loops to chromatin. This forms self-reinforcing loops where the methylation event (either histone methylation or DNA methylation) promotes siRNA generation, and the siRNAs in turn promote methylation (Holoach and Moazed 2015). Due to the stability of DNA methylation and the small loss of methylation in future cell divisions, epigenetic markers can be transmitted through many generations (Panchenko et al. 2016). See Figure 9 for an example of a self-reinforcing positive feedback loop.

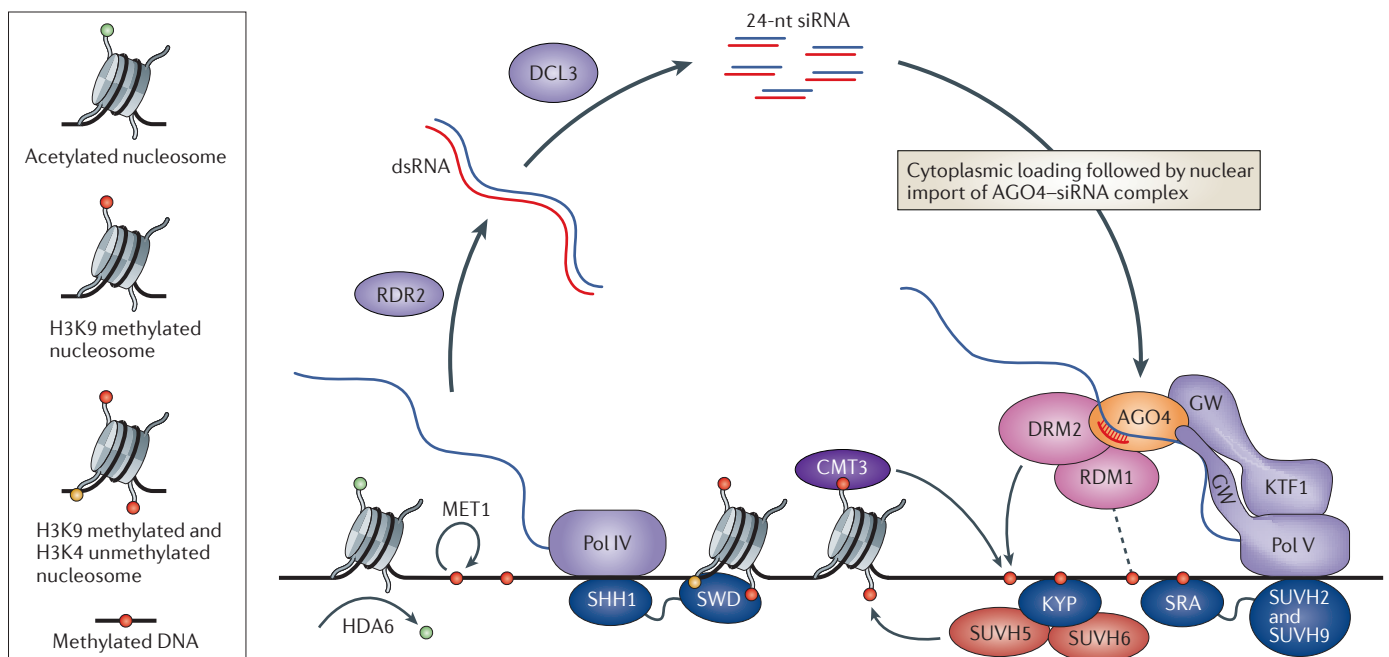


Figure 9 A self-reinforcing loop linking siRNAs to DNA and histone methylation in *Arabidopsis thaliana* “Elaborate feedback between small RNAs and DNA and histone methylation underlies a robust silencing pathway at sites of asymmetrical cytosine methylation in the *Arabidopsis thaliana* genome. Two plant-specific polymerases transcribe the critical RNAs. RNA polymerase IV (Pol IV) transcripts are processed by the RNA-dependent RNA polymerase RDR2 and the Dicer protein DCL3 into 24-nucleotide (nt) small interfering (siRNAs), while Pol V transcripts act as their targets. The Argonaute protein AGO4, the siRNA-dependent recruitment of which to Pol V transcripts is reinforced by interactions with the GW domains of Pol V and an associated elongation factor KTF1, in turn recruits the CHH DNA methyltransferase DRM2. RDM1 associates with the Pol V–AGO4–DRM2 complex and may link siRNA amplification to pre-existing DNA methylation. Meanwhile, another DNA methyltransferase that targets CHG sites for maintenance, CMT3, is recruited directly to methylated histone H3 lysine 9 (H3K9). Silencing by DNA methylation is augmented by H3K9 methylation, which is deposited by the enzymes KYP, SUVH5, and SUVH6. These methylation events are coupled to one another and to siRNA activity in several ways. KYP is recruited directly to methylated DNA, where it methylates neighboring histones, and the H3K9 methylation reader SHH1 recruits Pol IV to promote siRNA generation, while the DNA methylation readers SUVH2 and SUVH9 recruit Pol V to promote AGO4 targeting and further DNA methylation. Thus, the different methylation readers, RNA polymerases and AGO4 act together to create self-reinforcing interactions between pre-existing DNA methylation and siRNA amplification. Erasure of DNA methylation by mutations in either the histone deacetylase HDA6 or the maintenance DNA methyltransferase MET1 results in loss of siRNA biogenesis, emphasizing the importance of these self-enforcing interactions” (Holoach and Moazed 2015).

It is possible that RNAs are responsible for paternal transmission of epigenetic inheritance. In mammalian sperm, 10-20 fg (femtogram) of RNA are maintained in a mature sperm cell. Part of the RNA in sperm is made of fragments of coding transcripts, but there is also a variety of non-coding RNAs (approximately 24,000) found within the sperm (Kiani and

Rassoulzadegan 2013). Approximately 67% of the RNAs found in the sperm are enriched tRNA fragments. The number of these tRNA fragments decreases dramatically after fertilization, and some scientists believe that epigenetic modifications are being induced by these RNA molecules (Kiani and Rassoulzadegan 2013).

### *EPIGENETIC INHERITANCE*

Epigenetic changes, such as DNA methylation, that accrue throughout a person's lifetime can be inherited by that person's offspring (Heijmans et al. 2007; Johannes et al. 2009). This means that there can be variation in the phenotype without changing the genotype. If you vary the epigenome, the phenotypic variation increases while the genotypic variation remains the same.

Studies on plants have shown that inherited epigenetic profiles can remain stable across generations in the absence of natural selection (Johannes et al. 2009). The degree to which epigenetic variations are inherited is dependent on the location of the gene, the type of gene, and many other factors. Twin studies have shown that the DNA methylation of some genes are similar between twins due to inheritance of the methylation patterns, while DNA methylation patterns on other genes are dissimilar due to differing environmental conditions (Heijmans et al. 2007; Kaminsky et al. 2008). It should be noted that while most studies have found primarily a maternal transmission of epigenetics, it is also possible to have paternal transmission.

In studies on rats, it was found that endocrine disruptors could affect the epigenome in both females and males, while exposure of male rats to toxins prior to conception with a control female could also affect the epigenome of both male and female offspring (Cooney 2007). In other rat studies, it was found that paternal nutrition, toxin exposure, phenotypic variation, and paternal age can all lead to epigenetic variations in offspring (Curley et al. 2011). Different types

of paternal dietary conditions have been shown to have different effects in their offspring: fasting males have offspring with altered glucose metabolism; males with a high-fat diet have female offspring with altered pancreatic phenotypes; and males on a low-protein diet have offspring with decreased levels of cholesterol esters and altered expression of lipid/cholesterol biosynthesis genes (Rando 2012).

The genome of eukaryotes is organized as chromatin in the nucleus of the cell. The chromatin keeps track of all of the ongoing activities occurring in the cell, and contributes to the recruitment of transcription factors (Arnold et al. 2013). Histones are the proteins that package and order the DNA into nucleosomes, which are the structural units of chromatin. Histones are the primary protein found in chromatin, around which DNA winds (Backdahl et al. 2009). Small RNAs are key regulators of chromatin structure (Holoach and Moazed 2015). They have roles in RNA degradation and translational expression. They can modify chromatin.

A chemical modification to chromatin does not have to be transmitted to the next generation for it to be called epigenetic. It does, however, have to be transmitted across cell divisions to be considered epigenetic inheritance. The first definition of epigenetics was in 1942, and it states that epigenetics is “the casual interactions between genes and their products that allow for phenotypic expression” (Waddington 1942). As you can see, the original definition of epigenetics does not say anything about the modifications having to be transmitted to the next generation. While many definitions of epigenetics now mention heritability in association with epigenetics, not all of them do.

Epigenetic changes may be heritable through either meiosis or mitosis (Bossdorf et al. 2008). When epigenetics occurs within an individual, it occurs through mitosis. For example, if you have a highly methylated gene that divides, the subsequent cell will inherit this epigenetic



change. Throughout the cellular generations, the original epigenetic marks can be inherited. In somatic cells, epigenetic marks are mitotically inherited (Martos et al. 2015). Epigenetic marks may also be inherited through meiosis. The epigenomic profile of a mother may be passed on to her offspring *in utero*. Studies on plants have shown that inherited epigenetic profiles can remain stable across generations in the absence of natural selection (Johannes et al. 2009). The degree to which epigenetic variations are inherited is dependent on the location of the gene, the type of gene, and many other factors. Twin studies have shown that the DNA methylation of some genes are similar between twins due to inheritance of the methylation patterns, while DNA methylation patterns on other genes are dissimilar due to differing environmental conditions (Heijmans et al. 2007; Kaminsky et al. 2008). So, the answer to the original question is “no” and “no”. A chemical modification to chromatin does not have to be transmitted to the next generation for it to be called epigenetic, and it does not have to be transmitted through mitosis.

#### *EPIGENETICS AND DISEASE*

Monozygotic twins share identical DNA. If one twin develops a genetic disease, should the other twin not also develop the same disease? Studies on identical twins has found that one twin can have a genetic disease that the other twin does not have, or one twin has different apical levels of a disease than the other twin, or both twins develop the same genetic disease, but at different times (Hermus et al. 2007; Itin and Burger 2009). It is also not uncommon for monozygotic twins to have different diseases from the same family of disease. For example, both twins may have inflammatory bowel diseases, but one twin expresses this as Crohn’s disease, while the other twin expresses it as ulcerative colitis. What causes identical twins to not be identical? This discordance amongst monozygotic twins shows the importance of environmental factors in influencing both the onset and the severity of many diseases.

The methylation of a gene can be modified by environmental factors at all ages of a person's life (Barres and Zierath 2011; Brooks et al. 2010; Hayakawa et al. 2006; Itin and Burger 2009). This implies that a person's diet and exercise can affect that person's overall predisposition to many different diseases that are affected by methylation. As mentioned before, DNA methylation is mitotically stable. Due to this fact, it has long been assumed that environmental factors were unlikely to cause significant changes in the DNA methylation profiles of adults. Studies on monozygotic twins, however, have found that older pairs of twins have more divergent DNA methylation profiles than infant pairs of twins, indicating that environmental factors can significantly affect the epigenome throughout a person's life (Barres and Zierath 2011; Flintoft 2005). The older the sets of twins were, the more epigenetic differences there were between twins. Differences in DNA methylation profiles were also found to be greater between monozygotic twins who have developed a disease and between those twins who spent less of their life together (Brooks et al. 2010; Flintoft 2005). This once again shows that environmental factors can have a large impact on the DNA methylation of an individual. There was one case reported of monozygotic twins that had different Down syndrome phenotypes. The twins were aborted at 17 weeks of gestation, but postmortem analysis confirmed that the phenotypic differences seen on a sonogram (one twin had increased nuchal translucency and the other twin had a cervical cystic hygroma and a heart defect) were different Down syndrome phenotypes, even though the two had identical karyotypes (Grynberg et al. 2007). Differing epigenetic profiles is believed to be the cause of the different phenotypes of these identical twins. It is interesting that monozygotic twins can develop different epigenetic profiles while still within the womb and essentially sharing the same environment.

Many human diseases are multifactorial in origin, which means they are the result of interactions between the genetics of the individual and the environment in which he/she lives. The environment can directly affect the epigenetic profile of an individual, which in turn can affect the expression of that person's genetics. Thus, epigenetics may be the ultimate link between genetics and the environmental components that have been shown to play a role in many diseases, particularly autoimmune diseases (Backdahl 2010; Itin and Burger 2009; Lin et al. 2010).

Epigenetics may also explain differential susceptibility to diseases in different populations (Andraos et al. 2011). If a specific gene in one population is over-expressed or under-expressed due to an epigenetic cause, then the individuals in that population may be more or less likely to have a disease that is associated with the expression of that gene. Diseases with late-onset phenotypes, such as autoimmune diseases, “involve interactions between the epigenome, the genome, and the environment” (Feinberg 2007). There are two classes of epigenetic diseases – those diseases involving genes that are epigenetically regulated (such as imprinted genes) and those diseases that are caused by epigenetic changes that affect the entire epigenome (such as DNA methylation).

DNA methylation has been correlated with many different diseases, including, but not limited to: cancer, type II diabetes, arteriosclerosis, rheumatoid arthritis, and neurodegenerative diseases (Backdahl 2010; Backdahl et al. 2009; Heijmans et al. 2007). Cancer is characterized by global hypomethylation, gene-specific hypermethylation, and widespread chromatin modifications (Feinberg 2007). It is possible that early cancer detections can be made using epigenetics. One study found that methylation of the promoter region of the *CDKN2A* gene was found in the sputum of smokers up to three years prior to them being diagnosed with cancer

(Kristensen et al. 2008). Such an early warning system would allow an individual to begin cancer preventative treatments much earlier than usual. This would be similar to women who have mastectomies after they find out they have the breast cancer gene, even though they have not been diagnosed with cancer yet.

A study on patients with chronic kidney disease found that the amount of DNA methylation found in human peripheral blood cells correlates with varying degrees of inflammation as classified by levels of C Reactive Protein found in the patient's blood (Stenvinkel et al. 2007). It is generally accepted that the higher the levels of CRP in a person's blood, the more inflammation that person has in his/her body. Therefore, more methylation would mean more inflammation, most likely due to silencing of the methylated gene. Methylation at other loci on the human genome has been shown to correlate with other things, such as: age, gender, diet, and drug use (Backdahl et al. 2009). Therefore, it would seem that in time one could learn a lot about a person's lifestyle throughout their entire life just by looking at his/her epigenetic modifications.

It should be noted that it is not only cancer and autoimmune diseases that have been shown to correlate with DNA methylation. There are major psychotic disorders, including schizophrenia and bipolar disorders, which are also believed to have an epigenetic basis (Gavin et al. 2009; Kantlehner et al. 2011). It is currently unknown the origin and timing of these epigenetic changes during the duration of the psychosis. It is also not currently known what specific cell types in the brain have been affected by the methylated changes (Connor and Akbarian 2008). Postmortem brain studies of individuals with psychotic disorders have found alterations of mRNA levels that have been linked to inverse DNA methylation changes at the promoter regions (Connor and Akbarian 2008). More than 100 loci have been found to alter

CpG methylation in schizophrenic or bipolar disorder patients, most of which are gender-specific. One such locus is the peptidylprolyl isomerase E-like gene (PPIEL). It was found that individuals with bipolar disorder had significantly lower DNA methylation in the promoter region of this gene and significantly higher expression of this gene than was found in controls (Kuratomi et al. 2008). In addition to the previously mentioned diseases and disorders associated with epigenetics, the following diseases are also suspected of being influenced by DNA methylation: Alzheimer's disease, type II diabetes, atherosclerosis, lupus, and Parkinson's disease (Kantlehner et al. 2011).

Some scientists believe that a cure for many autoimmune diseases might lie in epigenetic regulation (Backdahl et al. 2009). If the original epigenetic profile of a gene has changed through environmental factors or the aging process (as is thought to be the case in many autoimmune diseases and cancers), then a “reset” of the epigenetic profile might result in curing the disease phenotype because many epigenetic modifications are reversible (Backdahl et al. 2009; Brooks et al. 2010). It has been shown in mammal studies that a deprivation of folic acid in the diet leads to changed DNA methylation levels in the liver that are then returned to normal upon the restoration of a diet with increased folic acid (Barres and Zierath 2011). This indicates that if the exact environmental triggers of a disease caused by hypermethylation are known, it might be possible to cure that disease through opposing environmental factors. Lab work has shown that certain chemical products (hydralazine and procainamide) can remove methyl groups from cytosines in CpG islands (Brooks et al. 2010). Hydralazine works to inhibit the pathway responsible for the induction of *DNMT1* and *DNMT3* transcription. Procainamide is a pharmaceutical antiarrhythmic agent used to treat cardiac arrhythmias that works by blocking open sodium channels.

## *EPIGENETICS, NUTRITION, AND THE ENVIRONMENT*

When studying epigenetics, nutrition is an important factor to consider. The foods that people eat are an important environmental regulator of the epigenome, and they can affect chronic inflammation, which is a hallmark of many chronic disorders seen in the world today (Barnett et al. 2010). For example, dietary fat has been shown to be a risk factor for the development of intestinal inflammation, such as that seen in individuals with Crohn's disease (Barnett et al. 2010). Individuals with these types of intestinal inflammation are told to try to maintain a low-fat diet to help offset the severity of the inflammation. Nutrients such as choline betaine, and folate have been shown to affect the biochemical mechanisms of DNA methylation regulation, and the deficiency of some nutrients can cause hypomethylation of the genome (Barnett et al. 2010). The methylation of the promoter region of some genes has been shown to be sensitive to levels of folate and selenium in the diet (Barnett et al. 2010), and this may prove to be an area in which a treatment program for some diseases may be discovered.

Many studies have shown a link between the nutrition of a pregnant female and the epigenome of her offspring. One study looked at the offspring of women who were pregnant during the Dutch famine of 1944 and found that there was a 5% reduction of DNA methylation of the *IGF2* gene's promoter region as compared to individuals who were not gestating at that time (Heijmans et al. 2008a). Further studies by the authors of that study found that about half of the promoters they examined showed methylation changes, usually with hypermethylation present in the famine group compared to a control group. Epidemiological studies in humans have linked maternal under-nutrition with increased risk of type II diabetes and obesity in children (Rando 2012). Studies have shown that women who had dietary constraints during gestation, but take folic acid during pregnancy or give their infants folic acid after weaning, give

birth to children that have altered phenotypes and epigenotypes than those that were originally present at conception (Burdge and Lillycrop 2010). One study has found that there is evidence of epigenetic transmission to subsequent generations. In this study, the grandmother rat had a protein-restricted diet and the mother rat had a normal diet. It was found that the grand-offspring were hypomethylated when compared to controls, even though their mothers had a normal diet (Burdge and Lillycrop 2010). If what is true in rats is also true in humans, then pregnant women should be especially careful in their consumptions since what they eat can affect not only their own offspring, but also affect subsequent generations of offspring.

The type of dietary restriction can have differing effects on one's epigenome. Caloric restriction up-regulates *SIRT1*, which is an enzyme required for cellular regulation, while protein restriction increases methylation of H3 histones and leads to gene silencing (Tollefsbol 2010). A high carbohydrate diet acetylates H3 histones, which activates transcription and DNA repair. Biotin deficiency activates histone acetylation and leads to dry skin, hair loss, rashes, and fungal infections. A high salt diet inhibits histone deacetylases and can lead to malfunctions in the heart. Zinc deficiency also inhibits histone deacetylases, and it can cause malfunctions in all tissues throughout the body (Tollefsbol 2010).

Nutrition is just one of many environmental conditions that can affect the epigenome. Other such conditions include (but are not limited to): smoking, toxin exposure, smog, parasite infections, drug use, and behavior (Cooney 2007). There are many environmental factors that can cause hypermethylation of a gene. Heavy metals are known to disrupt DNA methylation and chromatin (Feinberg 2007). While behavior may seem like a strange trait to pass on epigenetically, it has been proven to be the case in rats. Studies of rats have found that a daughter will nurse her offspring (high back with lots of licking or lowered back with little licking of

pups) in the same way in which her mother nursed her (Cooney 2007). This behavior was found to not be genetic in origin. In rats, it has been shown that epigenetics are important in regards to appearance, diabetes, obesity, overall health, behavior, and stress (Cooney 2007). If this is the case in rats and many other animal models, does it not follow that this could also be the case in humans? In one study on humans, it was found that differential methylation of the homeobox *DLX1* gene in monozygotic twins was associated with risk-taking behavior and a better ability to cope with anxiety (Kaminsky et al. 2008). They found that in twin pairs with differential methylation in this region, one twin was more likely to have a dangerous job (such as war correspondent), marry later in life, and have fewer children, while the other twin was more prone to a “safe” job, early marriage, more children, and higher overall levels of anxiety. The *DLX1* gene is involved in the formation and maintenance of interneurons in the hypothalamic pituitary adrenal axis, which is the stress center of the brain (Kaminsky et al. 2008).

#### *Epigenetics Research Questions*

When undertaking any type of research involving epigenetics, there are at least five questions that the researcher must keep in mind. First of all, how heritable are epigenetic patterns? It has been shown that the epigenetic patterns differ from gene to gene. The second question is: how much epigenetic variation is there? A small degree of variation in one gene could be significant while a large degree of variation in a different gene could be insignificant. The researcher must first determine what the normal levels of variation are for each gene he/she is studying. Third, how much do these differences contribute to phenotypic variation in a population/species? Fourth, how does natural selection act on epigenetic variation? The researcher must determine if natural selection is selecting for or against the variation or whether it is not acting on the variation at all. Natural selection increases fitness at both the individual and



population level due to epigenetic variation. Finally, how much of the variation in any phenotype is due to the environment, how much is due to genetics, and how much is due to epigenetic mechanisms? These are just some of the questions that must be considered when studying epigenetics.

### *Epigenetics of Crohn's Disease*

While few studies have looked at the epigenetics of Crohn's disease, many scientists agree that there is more than likely an epigenetic component to the disease (Backdahl 2010; Balasa et al. 2010; Barnett et al. 2010; Herrlinger et al. 2004; Petronis and Petroniene 2000; Risques et al. 2006). There are many environmental components that have been found to be associated with Crohn's disease. Among those environmental factors associated with the disease are cigarette smoking, non-steroidal anti-inflammatory drugs (NSAIDS), enteropathogens, and diet (Hubbard and Cadwell 2011; Kaser and Blumberg 2011). Cigarette smoking increases the formation of reactive nitrogen species (causing nitric oxide in the blood) and it also causes oxidation of plasma proteins and DNA (Backdahl et al. 2009). NSAIDS has been known to cause colonic bleeding and small intestinal inflammation in individuals with Crohn's disease, as well as cause a non-specific type of colitis (Klein and Eliakim 2010). There are enteric pathogens that can mimic Crohn's disease, as well as pathogens that can cause complications or relapse in individuals with the disease. For example, *Yersinia Enterocolitica* has been found in tissues of patients with long-standing Crohn's disease and may either precede or superinfect the disease (de Hertogh and Goeboes 2004).

It has been hypothesized that individuals with Crohn's disease have immune response genes that are highly methylated. This hypothesis is based five facts: (1) there are significant difference between maternal and paternal transmission, with maternal transmission being more

likely to occur than paternal transmission; there is a discordance of monozygotic twins affected by the disease; (3) there is impairment of the regulation of some cytokine genes (such as *TNF $\alpha$* ) in Crohn's patients, but the protein coding sequence is unaffected; (4) Crohn's disease affects a significantly higher percentage of females than males; and (5) epigenetic treatments have been shown to be highly effective in Crohn's disease patients (for example, glucosteroids increases the transcription of anti-inflammatory genes) (Petronis and Petroniene 2000). Glucosteroids are hormones produced in the adrenal glands that work to decrease inflammation. To date, very few studies have looked at the differential methylation of individuals with Crohn's disease. Of those studies that have looked at different methylation in Crohn's disease patients (Adams et al. 2014a; Backdahl 2010; Balasa et al. 2010; Cooke et al. 2012; Gonsky et al. 2014; Harris et al. 2012; Karatzas et al. 2014; Lin et al. 2010; McDermott et al. 2016; Nimmo et al. 2012), none have specifically studied the promoter regions of genes that affect autophagy in the intestines. Looking at differential methylation of genes is one way in which epigenetics – changes in gene expression that may or may not be heritable, but are not caused by changes in the DNA sequence – can be studied.

Most of the studies that have looked at methylation profiles in individuals with Crohn's disease have focused on the relationship between methylation and the risk of developing colitis-associated cancer, but it is also believed that methylation profiles could be used to diagnose Crohn's disease, as well as to determine the severity of the disease in an individual (Dabritz and Menheniott 2014). In order for this goal to be achieved, numerous studies on the methylation of different genes in Crohn's disease patients will need to be conducted.

During a chronic inflammatory state, two events can occur that can alter the methylation state of an individual's genes (Backdahl et al. 2009). The first thing that can occur is oxidative

stress, which is often found in the presence of chronic inflammation. The oxidation of methyl groups changes the conformation of 5-methylcytosine, which will lead to a loss of methylation after mitosis. The second thing that can occur is a build-up of reactive halogen compounds (a common by-product of inflammatory processes). The presence of these compounds can result in cytosine methylation mimicry where DNA methyltransferase *DNMT1* cannot distinguish between the true methylcytosine and the mimic halogenocytosine. *DNMT1* cannot separate the halogen bonds and so there is now a gain of methylation on the gene (Backdahl et al. 2009).

An inflammatory response to an infection is not necessarily caused by an epigenetic effect. Macrophages are large white blood cells (approximately 21 micrometers in diameter) that engulf and digest antigens through the process of phagocytosis (Karaiskos et al. 2011). They increase inflammation, stimulate the immune system, and can also decrease immune reactions by releasing cytokines. This is part of the normal immune response. In acute inflammation, there is an increase of leukocytes and plasma from the blood to the injured area (Bayarsaihan 2011). This is not due to an epigenetic effect.

Autoimmune diseases are characterized by the normal immune responses of the body being turned against its own tissues, which results in prolonged inflammation and subsequent tissue destruction. Some scientists believe that susceptibility genes interfere with self-tolerance and lead to the persistence of an abundance of autoreactive T-cells and B-cells; environmental factors then trigger cell/tissue injury and inflammation, thus activating the self-reactive lymphocytes to injure the tissues further; the final result is an autoimmune disease (Abbas et al. 2014). This can explain why individuals with the genetic susceptibility to a disease never show the disease phenotype, while also explaining why individuals exposed to the same environmental triggers may have different responses. Neither the gene nor the environment alone can claim sole

credit for the onset of an autoimmune disease; rather it is the interaction between the two that is responsible.

Autophagy is necessary in the mediation of the inflammatory response. “Autophagy degrades damaged organelles and proteins, in homeostasis and as a response to starvation, and is important for the clearance of pathogens (xenophagy), which is required for immunity to multiple different types of bacteria” (Cho and Brant 2011). Autophagy works by capturing organelles and cytoplasm and holding them within a membrane-bound organelle known as an autophagosome; once held, a lysosome comes along and destroys the microorganism within the autophagosome through the process xenophagy (Barrett and Chandra 2011). If this process is disrupted (through epigenetic or other means), the result is a production of inflammatory cytokines, which may lead to the poor response to intracellular bacteria that is found in individuals with Crohn’s disease (Barrett and Chandra 2011). Chronic inflammation, such as is seen in Crohn’s disease, is typically characterized by an invasion of undifferentiated CD4+ T-helper cells, known as Th0 cells (Barnett et al. 2010). The differentiation of TH1/TH2 cells from these Th0 cells is determined by epigenetic modifications of these cells (Backdahl 2010).

A study on patients with chronic kidney disease found that the amount of DNA methylation found in human peripheral blood cells correlates with varying degrees of inflammation as classified by levels of C Reactive Protein found in the patient’s blood (Stenvinkel et al. 2007). It is generally accepted that the higher the levels of CRP in a person’s blood, the more inflammation that person has in his/her body. Therefore, more methylation would mean more inflammation, most likely due to silencing of the methylated gene. When epigenetic effects alter gene expression to produce inflammation, you can say that there is an epigenetic effect of inflammation on the body. Epigenetic effects of inflammation are primarily

seen with chronic inflammation, rather than the normal inflammation you would see if you were to cut your finger, for example.

Epigenetic mechanisms can alter gene expression even in the absence of the original signal(s) that initiated them through a process known as metabolic memory (Intine et al. 2013). This is what is believed to happen in individuals with diabetes. During periods of hyperglycemia, a signal is transmitted to stabilize the individual's glycemic index. This signal causes epigenetic changes that are self-perpetuating even when the glycemic index has returned to normal, thus leading to the complications associated with diabetes (Intine et al. 2013). The epigenetic mechanisms in this case are responsible for altering the gene expression, rather than the gene being induced by a signal.

A study using bisulfite pyrosequencing showed that significant methylation differences in the epigenome can be determined from small sample sizes (Adams et al. 2014b). In this study, the authors used 18 pediatric case samples and 18 matched controls, and found 19 differently methylated regions, particularly the *HLA* region. A study on the methylation levels of *E. coli* in 10 Crohn's patients and 10 controls has shown that the amount of methylation is significantly correlated with the duration of the disease, as well as the stage of the disease (Pepoyan et al. 2015).

DNA damage repair occurs in the context of chromatin (O'Hagan 2014). Chromatin changes may occur around sites of DNA damage. DNA damage can occur in response to environmental exposures, as well as in response to genetic mutations. Chromatin is usually restored to normal as DNA damage repair is completed. In cases of prolonged exposure to toxins or in individuals with chronic inflammatory diseases, however, DNA damage-induced chromatin rearrangement can lead to permanent epigenetic alterations (O'Hagan 2014). These permanent

alterations may include reactivation of genes in the absence of a signaling molecule, as is what happens in the case of chronic inflammation.

Rheumatoid arthritis is a chronic inflammatory autoimmune disease. Studies have shown that epigenetic marks contribute to the pathogenesis of this disease (Bayarsaihan 2011). The inflammation seen in the disease is due to DNA methylation, histone acetylation, and miRNAs. Thus, this is an example of inflammation being the result of epigenetic mechanisms.

When an individual loses a limb through an accident, inflammation at the site will occur. This inflammation is acute inflammation, rather than chronic inflammation, and is the result of normal immune responses, rather than the result of epigenetic marks. No epigenetic changes have occurred in the somatic cells. Loss of limbs can occur through means other than accidents, however. Individuals with diabetes occasionally develop diabetic foot ulcers that result in the requirement of limb amputations (Rafehi et al. 2011). In these cases, there is often what is known as impaired diabetic wound healing. Investigations into epigenetic processes have found that epigenetic changes are likely responsible for impaired healing, as well as being key factors in diabetes and related complications (Rafehi et al. 2011). Thus, in the case of limb amputation due to diabetes, the inflammation present may be a result of epigenetic mechanisms, whereas in an accident it is not.

Crohn's disease often, but not always, occurs in multiple members of the same family. A recent study compared familial Crohn's disease with sporadic Crohn's disease (individuals who did not have a family history of the disease). This study found that in familial Crohn's disease, there is a statistically higher percentage of females (61%) that have the disease than in sporadic Crohn's disease (54% with  $p = 0.011$ ), statistically higher mother-to-daughter than mother-to-son transmission rate (36 vs 18,  $p = 0.02$ ), and a statistically higher mother-to-child than father-to-

child transmission rate (55 vs 32,  $p = 0.018$ ) (Zelinkova et al. 2012). This indicates that there is a sex-specific epigenetic inheritance pattern for Crohn's disease.

Studies have found that there is a familial risk of IBD – including both Crohn's disease and ulcerative colitis (Peeters et al. 1996), and there is a higher concordance rate for the disease in monozygotic twins than in dizygotic twins (Thompson et al. 1996). This concordance was low, with only 30% of the cases showing that when one monozygotic twin had Crohn's disease, the other twin also had the disease (Thompson et al. 1996). This low concordance indicates that more than a genetic component is at work in the disease phenotype.

To date, very few studies have looked at the differential methylation of individuals with Crohn's disease. Looking at differential methylation of genes is one way in which epigenetics – heritable changes in gene expression not caused by changes in the DNA sequence – can be studied. One study that has looked at the differential methylation of genes in individuals with Crohn's disease analyzed 1505 CpG sites of 807 genes in nine Crohn's disease patients and 26 control samples. The results showed differential methylation present in seven CpG sites (Lin et al. 2010). Of those sites, only three genes (*STAT5A*, *SERPINA5*, and *BGN*) showed consistently different methylation patterns in diseased tissues versus normal tissues (Lin et al. 2010). Due to the small sample size and the fact that many of the genes tested were not genes associated with Crohn's disease, additional work is recommended to determine whether there is differential methylation in the genes of Crohn's disease individuals. It should be noted that the three genes proposed for study in this project (*NOD2*, *ATG16L1*, and *PTPN2*) were not tested in the study conducted by Lin *et al.*

One recent study found over 3,000 sites in the genome that were differentially methylated in Crohn's patients compared to controls, while less than 1,500 sites were differentially

methyated between ulcerative colitis patients and controls with p values of  $< 0.05$  (McDermott et al. 2016). The most significantly differentially methylated genes were *ALOX5AP*, *HSPH1*, *CSPPI*, *ARFGEF1*, *SETD2*, *SERTAD2*, *SLC1A4*, *UPK3A*, *TRRAP*, *SMURF1*, *UPK3A*, *OPN3*, *NUP50*, *EXO1*, *SORL1*, and *BLID* (all of which had p values  $< 0.01$  that ranged from  $1.20e^{-20}$  to  $.49e^{-22}$ ).

A study on peripheral blood leukocytes using DNA methylation microarrays identified a single association with IBD at the testis, prostate, and placenta-expressed protein (*TEPP*) when DNA isolated selectively from peripheral blood cells was analyzed (Harris et al. 2012). This study used two different high-throughput microarray-based methods to analyze genome-wide DNA methylation, and the results were validated with bisulfite pyrosequencing. There was a 96.3% correlation between using microarrays and bisulfite pyrosequencing to test for DNA methylation. This correlation was statistically significant with  $P < 0.00001$  (Harris et al. 2012). This study primarily used monozygotic twins that were discordant for IBD. While most genes in this study were not found to be associated with IBD, the authors believe that this could be due either to the type of tissue used or due to discordant monozygotic twins being epigenetically less different from the healthy twin sibling at the pathogenic loci tested, as compared to non-related patients and controls (Harris et al. 2012).

One study looked at whether there were methylation differences in individuals with Crohn's disease versus ulcerative colitis versus healthy individuals. This study found that Crohn's disease and ulcerative colitis patients were more likely to have differential methylation for the following genes when compared to controls: *CARD9*, *ICAM3*, *IL8RB*, *THRAP2*, *FANCC*, *GBGT1*, *DOK2*, *TNFSF4*, *TNFSF12*, and *FUT7* (Cooke et al. 2012). They did not find any differences between individuals with the two forms of irritable bowel disease (Crohn's disease



and ulcerative colitis) in this study. Many of these genes had previously been identified as being highly correlated with irritable bowel disease.

A study by Gonsky *et al.* found that there was differential methylation in the *IFNG* gene (that codes for interferon gamma - *IFN- $\gamma$* ) in individuals with Crohn's disease. This differential methylation leads to decreased function of the *IFN- $\gamma$*  gene, which is needed for the proper working of both the innate and adaptive immune response (Gonsky et al. 2014). It is logical that a decrease in the function of *IFN- $\gamma$*  would lead to problems with autoimmune diseases, Crohn's disease just being one example of many.

One study found an association between DNA-methylated B cells and individuals with Crohn's disease. This study used peripheral blood of eighteen individuals (9 with Crohn's disease and 9 with ulcerative colitis) to determine differences in B cell line methylation using microarray techniques (Lin et al. 2012). They identified 11 IBD-associated CpG sites, 14 Crohn's disease-specific CpG sites, and 24 ulcerative colitis-specific CpG sites with methylation changes in B cells (Figure 10). Many of the genes are involved in immune and inflammatory response functions.

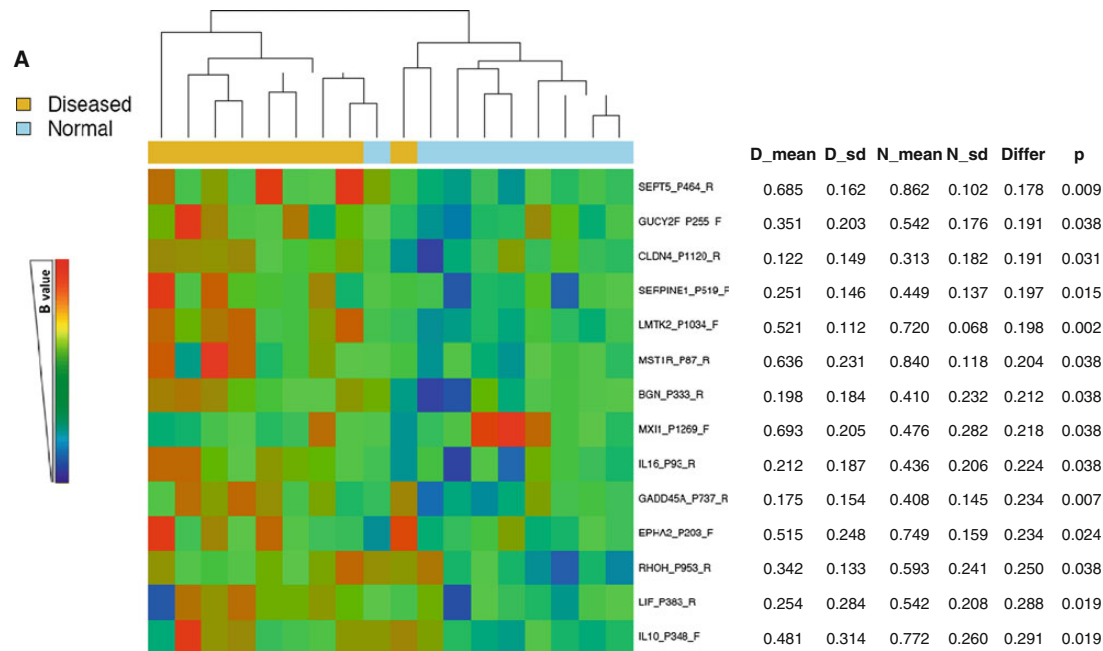


Figure 10 Cluster analysis of 14 Crohn's disease-associated methylation CpG sites in B cell lines. Left panel Hierarchical clustering heat map of methylation data. Right panel Data analysis of the 14 CD-associated loci from B cells of IBD (CD and UC) patients and non-IBD siblings. Gene symbols contained within the target ID before the first underscore. Rows correspond to CpG sites. Color indication for samples: orange represents B cells from IBD patients, and light blue represents non-IBD siblings; rows correspond to CpG sites; b-value: red indicates increased methylation, and blue indicates decreased methylation. Right panel: D\_mean = mean b-value (fractional methylation from 0 to 1) of diseased tissue; D\_sd = standard deviation of diseased tissue; N\_mean = mean b-value of normal tissue; N\_sd = standard deviation of diseased tissue; Differ = absolute mean difference between diseased and normal tissues. p = p value (Lin et al. 2012). The heat map on the left shows that there is increased methylation in individuals with IBD for most of the genes tested, whereas their healthy siblings primarily showed a decrease in methylation.

There have been numerous association studies that have found genes purported to be associated with Crohn's disease. Later studies have been unable to reproduce many of these findings. The difficulty in reproducibility may be due to the fact that some type of regulation other than genetic is contributing to the phenotype of the disease. According to one scientist, "methylation of DNA represents the only flexible genomic mechanism that provides a link between the predetermined genetic basis of disease susceptibility and environmental component and the resulting disease phenotype" (Backdahl 2010). One of the most comprehensive studies on the epigenetics of Crohn's disease analyzed the methylation status of 27,578 CpG sites across the genome in 21 adult females with Crohn's disease and 19 adult female controls (Nimmo et al.

2012). This study found significant differential methylation in genes involving immune activation, immune response, defense responses to bacteria, and immune system processes. They found that methylation status was predictive of disease status with a sensitivity of 0.71 and a specificity of 0.83 (Nimmo et al. 2012). For medical tests, sensitivity and specificity results above 0.90 are considered to be highly credible (Pewsnar et al. 2004). The following genes were found to be significantly hypomethylated in Crohn's patients: *CD22*, *CEBP3*, *ADPH*, *SP11*, *SSBP2*, *FLJ22746*, *RABGEF1*, *IL21R*, *RNASE3*, *CASP10*, *C10or91*, *IL19*, *CTSG*, *SFTPD*, *ACOT11*, *TLR8*, *FYP*, *ICAM2*, *SLC5A8*, *CD300LF*, *GL4X2*, *RNASE2*, *SH3TC1*, *CD59*, *YIPF4*, *SH3TC2*, and *PLEKHA6*; the following genes were found to be significantly hypermethylated in Crohn's patients: *APBA2*, *MAPK13*, *CCDC47*, *ABI3*, *SLAMF7*, *PRF1*, and *FASLG* (Nimmo et al. 2012). This study did not look at the *NOD2*, *ATG16LI*, or *PTPN2* genes.

Since Crohn's disease is characterized by chronic inflammation, studying the methylation profiles of individuals with the disease is a particularly well-suited approach. Many immune system processes that cause inflammation are regulated by epigenetics, and therefore many autoimmune diseases may be epigenetic in origin (Backdahl 2010). Chronic inflammation, such as is seen in Crohn's disease, is typically characterized by an invasion of undifferentiated CD4+ T-helper cells, known as Th0 cells (Barnett et al. 2010). The differentiation of TH1/TH2 cells from these Th0 cells is determined by epigenetic modifications of these cells (Backdahl 2010). Demethylation of the *Foxp3* promoter of regulatory T-cells is characteristic of this type of cell. B-cells and cells involved in innate immunity also have differences in their epigenetic profiles (Backdahl 2010).

One of the studies that have looked at the epigenetics of Crohn's disease looked at the interferon regulatory factor 5 (*IRF5*) gene (Balasa et al. 2010). There is a SNP in the promoter

region of this gene that affects the CpG dinucleotide-dense genomic region. This polymorphism is associated with Inflammatory Bowel Disease (IBD). The authors believed that individuals with this polymorphism might have a different methylation profile than individuals without the polymorphism. Their results, however, indicated that there was no significant differences in methylation, and thus the differential methylation of the *IRF5* promoter is unlikely to be a cause of IBD (Balasa et al. 2010). While these study yielded negative results, the authors say that the study shows the need for large-scale epigenomic studies due to the fact that candidate gene approaches are less likely to yield positive results than epigenetic correlate studies of inflammatory bowel diseases (Balasa et al. 2010).

Another study looked at genetic and epigenetic changes in Signal Transducer and Activator of Transcription 4 (*STAT4*), a gene whose elevated expression is associated with Crohn's disease (Kim et al. 2012). *STAT4* is required for the development of Th1 cells from naïve CD4+ T cells, as well as production of *IFN- $\gamma$*  in response to *IL-12*. This study used blood and colon mucosa, and the mRNA and DNA was isolated to test for methylation using real-time PCR. The researchers found that *STAT4* had elevated expression in Crohn's disease patients, and there was a significant correlation between risk alleles and methylation of the promoter region (Kim et al. 2012) They concluded that the DNA methylation of the gene is associated with genetic polymorphisms, which provides insight into the interaction between genetic and epigenetic factors that contribute to the development of Crohn's disease. For a look at the current research on differential methylation of genes in individuals with Crohn's disease, see Table 8.

Table 8 A summary of the current research on differential methylation in individuals with Crohn's disease, as compared to controls. Table includes the authors of the paper, the number of Crohn's disease patients, ulcerative colitis, and controls used in the study. Also included is the type of sample collected, the genes with differentially methylated loci, and the number of loci found to have methylation.

Authors	Subjects	Samples	Highly differentially methylated loci	Number of loci with methylation
(Harris et al. 2012)	18 CD, 15 UC	Peripheral leukocytes PBMCs	<i>TEPP</i>	1
(Lin et al. 2012)	9 CD, 9 UC	Epstein-Barr virus-transformed B cells	<i>Bcl3, PPARG, STAT3, OSM, STAT5, IL12RB, SOX1, COL18A1</i>	49
(Nimmo et al. 2012)	21 CD, 19 controls	Whole blood	<i>MAPK13, FASLG, PRF1, S100A13, RIPK3, 8L-21R</i>	50
(Hasler et al. 2012)	20 UC, 135 controls	Intestinal biopsy specimens (whole tissue)	<i>CF1, SPINK4, THY1/CD90</i>	61
(Cooke et al. 2012)	16 CD, 16 UC, 8 controls	Rectal biopsy specimens (whole tissue)	<i>THRAP2, FANCC, TNFSF4, TNFSF12, FUT7, CARD9, ICAM3, IL8RB</i>	>500
(Lin et al. 2010)	9 CD, 17 UC, 26 controls	Intestinal tissue from surgery (whole tissue)	<i>BGN, SERPINA, TNFSF1A, AATK, GABRA5, MAPK10, STAT5A</i>	7
(Adams et al. 2014a)	18 CD, 18 controls (all pediatric patients)	Peripheral blood leukocytes	<i>HLA, MIR21</i>	19
(Gonsky et al. 2014)	17 CD, 39 controls	Peripheral blood leukocytes	<i>IFNG</i>	1
(Kim et al. 2012)	9 CD, 9 UC, 29 controls	Peripheral blood leukocytes and colonic mucosa	<i>STAT4</i>	1

One reason why it is important to study the epigenetics of diseases such as Crohn's disease is that unlike genetic variations, epigenetic modifications are potentially reversible

(Barnett et al. 2010). If it is found that epigenetic modifications are responsible for the Crohn's disease phenotype, then a permanent cure could be developed reversing the epigenetic modifications.

Many studies have found an association between Crohn's disease and the *NOD2* gene. This association implies that the *NOD2* gene is not functioning properly in individuals with the disease. Although never tested, it has been hypothesized that in addition to the SNPs of the *NOD2* gene, the epigenetic mechanisms of the gene may be one way in which the gene function is regulated (Barnett et al. 2010). Testing if there is differential methylation of this gene between healthy and diseased individuals may be important in determining one of the causations of Crohn's disease.

#### *Crohn's disease treatments and future prognosis that are affected by methylation*

Thiopurine drugs, such as azathioprine and mercaptopurine, have long been used to treat patients with Crohn's disease with an efficacy of 50-80% (Herrlinger et al. 2004). Thiopurine drugs are purine antimetabolites that are used for treating many autoimmune disorders (Sahasranaman et al. 2008). In order to function, thiopurines require bioactivation by hypoxanthine phosphoribosyltransferase, which turns them into either thioinosine monophosphate or thioguanosine monophosphate nucleotides (Herrlinger et al. 2004). Thiopurines can be inactivated by the catalyzation of S-methylation thiopurine by thiopurine S-methyltransferase (*TPMT*) (Herrlinger et al. 2004). *TPMT* is inversely related to the accumulation of cellular 6-thioguanine nucleotides (6-TGN), which has been shown to be methylated in individuals with Crohn's disease (Herrlinger et al. 2004).

Many individuals with Crohn's disease often progress to colorectal cancer over time. Recent studies have found that this progression may be due to "chronic cycles of injury,

inflammation, repair and telomere shortening” (Risques et al. 2006). Methylation also plays a role in the progression to colorectal cancer, as chronic inflammation has been linked to activation of the *TP53* gene that works as a tumor suppressor in preventing cancer (Risques et al. 2006). It has been found that low levels of selenium and folate in the mother’s diet during gestation and lactation can cause plasticity of the epigenetic modifications that may eventually lead to colon cancer, and this may also relate to the epigenetic regulation of IBD in that individual (Barnett et al. 2010).

As has been shown thus far, Crohn’s disease is characterized by chronic inflammation. This inflammation may be affected by epigenetic modifications such as methylation. The diet of an individual with Crohn’s disease can greatly affect the phenotype of the disease. For many individuals with IBD, certain foods (spicy, containing gluten, high in fiber, or high in fat) can exacerbate the condition. More studies need to be performed to completely understand the “complex interplay between nutrients, epigenetic events and chronic inflammation” that is seen in individuals with Crohn’s disease (Barnett et al. 2010).

Scientists have recently been attempting to use genetic and epigenetic markers to predict the efficacy in Crohn’ disease therapeutic agents, and to avoid toxicity (Gabbani et al. 2017), but so far there have been few successful attempts to tailor a therapy to an individual based on their genetic markers. In addition to looking at markers in disease treatments, scientists have also recently been studying the epigenetic modifications of various pathways that are known to lead to inflammation in individuals with IBD. For example, a recent study has found seven differentially expressed epigenetic modulators, particularly the down-regulation of lysine acetyltransferase 2B mRNA (*KAT2B*), which has been shown to reduce the production of IL-10,

a crucial anti-inflammatory cytokine that has decreased expression in individuals with IBD (Bai et al. 2017).



## CHAPTER 3: MATERIALS & METHODS

The purpose of this study was to determine the extent to which genetic, epigenetic, behavioral, and environmental factors contribute to the Crohn's Disease phenotype. There were five steps that were taken in order to assess the impact of each of these factors. The first step was the collection of data. This included both administering a questionnaire to the participants, as well as taking samples of their genetic material. The second step was to extract and purify the DNA from the samples taken. The third step was to bisulfite convert the samples. The fourth step was to determine whether or not there was differential methylation within the regions tested. The fifth (and final) step was to completely analyze the data collected using statistical methods. This section will fully describe each of the steps taken during this study.

For this project, I collected genetic material, medical histories, and environmental variables from individuals with Crohn's disease, individuals with ulcerative colitis, and individuals that are phenotypically normal. Crohn's disease and ulcerative colitis both involve the digestive tract, yet each autoimmune disease affects it in different ways. This project will be used as an attempt to answer the questions of why ulcerative colitis only affects the colon, while Crohn's disease can affect the entire digestive tract and why antibodies only attack certain parts of the system. The primary questions that will be asked in my dissertation will be: Do Crohn's disease and ulcerative colitis have the same cause? Is it epigenetic regulation that influences which disease phenotype the individual has? If it is epigenetic, are there differences in gene methylation that trigger the disease(s)? Or is the disease triggered by environmental factors? By answering these questions, this project will add valuable information to science and will help our understanding of these diseases. It is also hoped that this project may lead to a non-invasive way of diagnosing Crohn's disease.

In order to determine if gene methylation can be used as a diagnostic tool for Crohn's disease, this project compares the methylation status of diagnosed Crohn's patients with individuals that are phenotypically normal. DNA and information relevant to this project (via a questionnaire attached to the end of this proposal) were collected from individuals with Crohn's disease, individuals with ulcerative colitis, and phenotypically normal individuals for purposes of comparison.

This project was approved by the IRBs of the University of Kansas and the University of Kansas Medical Center. Participants were recruited for this project at the offices of Dr. John Bonino, a gastroenterologist at Kansas Medical Center. Some of these individuals had Crohn's disease, some had ulcerative colitis, and others were phenotypically normal. The individuals were given an informed consent form to read and sign prior to participation. That form can be found in Appendix 1.

## ***SUBJECT SELECTION CRITERIA AND SAMPLE SIZE JUSTIFICATION***

Buccal swabs and information relevant to this project (via a questionnaire – see Appendix 2) was collected from individuals with Crohn’s Disease, as well as individuals without the disease for purposes of comparison. The questions are designed to elucidate information regarding the overall health of the individual, as well as possible behavioral and/or environmental exposures that could have an effect on Crohn’s Disease. General questions were asked regarding the age, gender, and ethnicity of the participant. Specific questions were asked pertaining to the health of the individual, such as whether or not the individual had ever had certain diseases, high blood pressure, or high cholesterol levels. The behavior of the individual was based on questions relating to smoking, alcohol consumption, caffeine intake, allergies, and stress at the time of diagnosis. The environment of the individual was determined based on questions relating to exposure to different environmental chemicals and substances, as well as whether the individual has primarily lived in a rural or urban environment.

The questionnaire that was administered was developed using the PhenX Toolkit.

According to [phenxtoolkit.org](http://phenxtoolkit.org),

“The PhenX Toolkit offers well-established, broadly validated measures of phenotypes and exposures relevant to investigators in human genomics, epidemiology, and biomedical research. The measures in the Toolkit are selected by **Working Groups** of domain experts using a consensus process. The Toolkit provides detailed protocols, information about the measures, and tools to help investigators incorporate PhenX measures into their studies. Inclusion of PhenX measures facilitates cross-study analysis downstream, thus increasing the scientific impact of each individual study.”

What this means is that the data collected for this project can be compared to other studies that have previously been reported, as well as be available for comparisons in future studies. This information was collected at the University of Kansas Medical Center under the direction of Dr. John Bonino, a gastroenterologist.

The subjects were chosen on a voluntary basis. Power and Precision<sup>TM</sup> software says that this study will have power of 95% to show that the mean for methylation in Non-Crohn's disease individuals is the same (neither lower nor higher) as the mean for methylation in Crohn's disease individuals. This assumes that the means for the Crohn's disease and Non-Crohn's disease populations are precisely equal with a common within-group standard deviation of 1.00, that a difference of 0.47 points or less is unimportant, that the sample size in the two groups will each have 25+ individuals, and that the alpha is set at .05. According to Balasa *et al.* (2010), significant differences in methylation of more than 5% can be obtained from a minimum of 10 samples in each group.

Formally, the null hypothesis is that the mean for methylation in Non-Crohn's disease individuals is not 0.47 points lower or higher than the mean for methylation in Crohn's disease individuals, and the study has power of 95.1% to reject this null. Equivalently, the likelihood is 95.2% that the 95.0% confidence interval for the mean difference will exclude a difference of 0.47 points in either direction.

### ***DNA EXTRACTION FROM BUCCAL SWABS***

Genetic and epigenetic material was obtained by taking saliva from the individuals. While saliva, by itself, does not contain DNA, it does contain epithelial cells from the cheek that has DNA present in them. Due to the fact that Crohn's disease affects an individual's entire gastrointestinal tract from the mouth to the anus, the use of buccal swabs for DNA collection is an acceptable practice (Barrett and Chandra 2011). DNA was extracted from the buccal swabs using the QuickExtract™ DNA Extraction Solution manufactured by Epicentre. This solution is used to quickly and efficiently extract PCR-ready genomic DNA from any tissue type. To perform the extraction, the swab containing buccal cells is added to a microcentrifuge tube filled with QuickExtract™ solution. The tube is then heated at 65°C for 6 minutes and 98°C for 2 minutes. Once the procedure is complete, the microcentrifuge contains PCR-ready DNA.

## **PRIMER DESIGN**

To avoid PCR bias in the bisulfite conversion to real-time PCR step, I designed primers based on the current literature (Wojdacz et al. 2008): the fewer CpG sites, the better; keep CpG sites away from the 3' end of the primer; melt temperature ( $T_m$ ) of primer should be around 65°C to run PCR annealing step at 60°C; include at least one thymine near the 3' end that is away from a CpG site; check primer for primer dimer complications. In addition, the primer should have at least two CpG sites, and non-CpG cytosines should be included in the sequences (Kristensen et al. 2008). No single nucleotide polymorphisms should be found in the amplified portion of DNA. This was checked using BLAST searching of the SNP database. The reason that SNPs cannot be in the amplicon is that they interfere with the melting profile if found between primers (Kristensen et al. 2008). The primers will be used on bisulfite-converted samples.

Here is a step-by-step of how the primers were designed:

1. Go to the website <http://ncbi.nlm.nih.gov>
2. Search for the gene (ex: *NOD2*) and specify *Homo sapiens*
3. Click on one of the sequences (it gives GenBank number and more information including the source of the sample and who uploaded it into GenBank)
4. On the right, click on Pick Primers (you want one with a PCR product of 100-150)
  - a. Make sure there is a CpG near the 3' end of the primer, as well as other cytosines not in CpG being present in the primer
  - b. There should be no SNPs in the resulting amplicon between primers; do a BLAST searching of the SNP database to make sure (dbSNP Build 127)
5. Take your primer (copy forward) and go to <http://idtdna.com>
6. Click on SciTools → OligoAnalyzer
7. Paste in the sequence
8. Click Analyze (gives GC content,  $T_m$ , etc..) and check for the following
  - a. Hairpin: you do not want a hairpin above 40°C
  - b. Self-dimer: the  $\Delta G$  should be more positive than -7, if possible
9. Copy the reverse sequence and put it in, then calculate
  - a. Heterodimer: needs to be more positive than  $\Delta G$  of -9 (the first one listed is always the most energy)
10. BLAST it on GenBank
  - a. Put in your forward and click “run blast”. You want the result to come back and say 100% in the 5<sup>th</sup> column for your species/gene

b. Repeat for your reverse.

Table 9 The primer sequences for the three genes following bisulfite conversion. The CpG sites are in bold, and the converted cytosines are listed as a capital T in the forward primer and a capital A in the reverse primer.

Gene	Primer sequences (CpG sites in bold and converted Cs as capital Ts or As)	Annealing temp (°C)	Amplicon size (bp)
<i>NOD2</i>	F – <b>acgtggTttgggc</b> R – acaccccaActaAtA <b>acg</b>	48.6 49.1	65
<i>ATG16L1</i>	F – TTT <b>tcgTtcg</b> Tattggtggc R – <b>cg</b> AAAAaaAtca <b>AcgAcg</b>	53.6 47.3	81
<i>PTPN2</i>	F – <b>cg</b> TtaTTT <b>acgga</b> TTa <b>Tc</b> R – Ac <b>ccgt</b> AA <b>tc</b> ccct <b>Ac</b>	42.1 46.3	87

## ***METHODS FOR DETECTING DNA METHYLATION***

Sections of DNA can be methylated by the transfer of a methyl group from an S-adenosyl methionine to the 5<sup>th</sup> position of the pyrimidine ring of a cytosine. When a gene is highly methylated, the expression of the gene is usually either reduced or completely silenced (Bird et al. 1995). The reduction of expression or silencing of a gene can cause or amplify diseases. Looking at differential methylation of genes is one way in which epigenetics – heritable changes in gene expression not caused by changes in the DNA sequence – can be studied (Bird 2007).

Many methods have been used to detect DNA methylation in mammalian cells. Some scientists have used restriction endonuclease-based analysis. The most common pairs of enzymes used are *HpaII-MspI* (recognizes CCGG), *SmaI-XmaI* (recognizes CCCGGG), and *McrBC* (cleaves DNA containing methyl-cytosine) (Gupta et al. 2010). This method is limited by the availability of restriction enzyme sites present in the targeted DNA. Immunoprecipitation based methods have also been used. Specific antibodies can be used to identify methylated cytosines and the proteins that bind to them (Gupta et al. 2010). These methods, unlike bisulfite conversion methods, do not provide DNA methylation information for a single-nucleotide position. Mass spectrometry-based methods have also been used, however this is not the method of choice due to limited throughput and the high cost for performing DNA methylation analysis (Gupta et al. 2010). Due to the previously mentioned reasons, the bisulfite conversion method is considered to be the best method to use.

### ***Bisulfite Conversion***

The most common method of testing for methylation is currently the bisulfite conversion method. In this method, all of the non-methylated cytosines within a region are converted to



uracil (Heijmans et al. 2008b). With this method, it is possible to determine the extent of methylated cytosines within each region of interest. While many other methods have been used in the past to detect DNA methylation, bisulfite conversion is now considered the “gold standard” (Gupta et al. 2010). Bisulfite conversion is used to detect 5-methylcytosines in DNA. The chemistry for bisulfite conversion was first developed in 1973 during treatment of calf thymus DNA with bisulfite (Shapiro et al. 1973). The bisulfite conversion method was first used as a means to detect differential methylation patterns in humans in 1992 (Frommer et al. 1992). During bisulfite conversion, unmethylated cytosines are converted to uracil, while methylated cytosines remain unaltered (Barnett et al. 2010; Callinan and Feinberg 2006; Reece 2012). Subsequent PCR amplification then shows the methylated cytosine bases as thymine, which can be compared to the original sequence for the presence of methylation. Even though this is the gold standard for testing methylation profiles, there are still some difficulties with the method. When the unmethylated cytosines are converted to uracil, sometimes not all of them are converted. It has been found that there may be anywhere from 0-2% of unconverted unmethylated cytosines in a sample (Kristensen et al. 2008). This can lead to overestimation of methylation levels when looking at what is actually low levels of methylation.

For this project, bisulfite conversion was performed using the EZ DNA Methylation-Lightning<sup>TM</sup> Kit manufactured by Zymo Research. This kit was chosen as it has one of the highest correct conversion rates (>99.5%), one of the fastest conversion rates (1 hour), and the lowest minimum starting DNA required (0.1 ng). This project used less than the optimal amplicon (150-300bp). The reason for this was that in order to meet all of the requirements for the primer design, a larger amplicon in the promoter region could not be found. The manufacturer’s instructions say that 35-40 cycles are needed for successful PCR amplification,

and that annealing temperatures between 55-60°C typically work well. Kristensen states that the annealing temperature should be based on the primer. Melt temperatures are lower for bisulfite-converted primers than they are for non-converted primers. For this project, the *NOD2* and *PTPN2* primers require annealing temperatures of 45-50°C, rather than what the manufacturer recommends for bisulfite conversion.

Once bisulfite conversion has been applied to the sample, a variety of methods can be used to measure the amount of methylation present in the sample. The methods available include: pyrosequencing, CE-based sequencing, PCR sequencing, methylation-specific PCR, and single strand conformational polymorphism analysis (Barnett et al. 2010). Once all of the non-methylated cytosines were converted to thymines, a melting-curve analysis was performed using real-time PCR. The bisulfite-converted DNA was amplified using PCR and the quantity of the product was analyzed. This method allows one to determine the ratio of methylated to unmethylated products by comparing the different peaks generated in the melting-curve analysis, and is useful for detecting even low-level methylation (Kristensen et al. 2008).

Real-time polymerase chain reaction is a molecular biology technique that monitors the amplification of a targeted molecule of DNA in real time. Real-time PCR is performed in a thermocycler that is able to illuminate each sample. The thermocycler heats and cools the samples in cycles. There are three stages of PCR. In the first stage, the thermocycler is heated in order to separate the double-stranded DNA; in the second stage, the temperature is lowered and then raised slightly in cycles in order to allow binding of the sequence-specific primers to the DNA template; in the final stage, the temperature is again raised so that polymerization occurs through the use of a polymerase enzyme. Real-time PCR is a semi-quantitative method that can be used to estimate percentages of methylation, as well as the location of the methylation (Reece

2012). By using higher temperatures during the holding stage at the end of each cycle on real-time PCR, the rate of false-positives from unconverted cytosines decreases drastically (Kristensen et al. 2008).

### ***POLYMERASE CHAIN REACTION (PCR)***

When Real-Time PCR is performed on bisulfite converted samples, methylated DNA is preferentially amplified over unmethylated DNA with a 2.3°C-5°C difference in average melt temperatures between the unmethylated and the methylated samples (Reece et al. 2012). (Reece et al. 2012). If there is significantly different methylation, one would expect the methylated samples to have a lower  $C_T$  and a higher  $T_M$  than non-methylated samples. Those samples that are unconverted have a slightly higher melt temperature than the methylated samples.

In order to perform real-time PCR, the following quantities were added to each well of the MicroAmp<sup>TM</sup> Fast Optical 48-well reaction plate:

10 µl SYBR® Green PCR Master Mix (SYBR® Green I Dye; AmpliTaq Gold® DNA Polymerase, UP; dNTPs; passive reference; optimized buffer components)  
0.4 µl Forward primer  
0.4 µl Reverse primer  
4.2 µl Molecular water  
5 µl Sample

According to the StepOne<sup>TM</sup> user's manual, PCR should be performed at: 50°C for 2 minutes; 95°C for 10 minutes; 40 cycles of 95°C for 15 sec (Sato et al.) then 60°C for 1 min (anneal/extend). Per Kristensen et al, the PCR should be set to: 95°C for 15 minutes for enzyme activation, followed by 45 cycles of 95°C for 20s, annealing at the appropriate temperature for 30s (based on the primer), 72°C for 30s, and one cycle of 95°C for 1 minute. The appropriate annealing temperature is 5°C lower than the melt temperature. For this project, the PCR was performed at:

1. 95°C for 15 minutes
2. 45 cycles of
  - a. 95°C for 20s
  - b. Annealing temperatures should be 5 degrees lower than melt temp so:
    - i. *NOD2bc*: 43°C for 30 seconds
    - ii. *ATG16L1bc*: 42°C for 30 seconds

- iii. *PTPN2bc*: 37°C for 30 seconds
- c. 72°C for 30sec
- 3. 1 cycle of 95°C for 1 minute

The standard deviation of bisulfite pyrosequencing of a single CpG site has been found to vary between 1% and 5% (Balasa et al. 2010). Methylation differences of greater than 5% for a minimum of 10 samples are considered significant differences (Balasa et al. 2010).

### ***DATABASE***

All data (environmental, genetic, and epigenetic) was entered into a database and analyzed using the statistical programs SPSS24 and Minitab17. The non-numerical variables were given a value (i.e. 1=yes; 2=no) for analysis.

## STATISTICAL TESTS

After entering all data into the statistical program, SPSS, multiple statistical tests were run. The following tests were used: correlations and non-paired t-test. For a list of the counts for each variable from the questionnaire, please see Appendix 3. The chart in Appendix 3 lists the number of individuals with each variable who have Crohn's disease, who have ulcerative colitis, and who have neither.

The first step taken during the statistical analysis phase was to determine whether or not there was an association between Crohn's disease and any of the variables. A correlation measures the strength of the relationship between two variables. Just because two variables are correlated, however, does not necessarily mean that one variable is the causation of the second variable (correlation does not equal causation). A Pearson's  $r$  correlation analysis was performed on each variable. The population correlation coefficient  $\rho_{X,Y}$  between two random variables  $X$  and  $Y$  with expected values  $\mu_X$  and  $\mu_Y$  and standard deviations  $\sigma_X$  and  $\sigma_Y$  is defined as:

$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y},$$

where  $E$  is the expected value operator,  $\text{cov}$  means covariance, and,  $\text{corr}$  a widely used alternative notation for Pearson's correlation. This test can determine whether or not there is a relationship between any two variables.

While a Pearson's correlation works for continuous variables, for nominal variables of sample sizes less than 1000 it is best to use Fisher's exact test (McDonald 2014). Fisher's exact test uses a 2x2 table to test the null hypothesis that the relative proportions of one variable are independent of a second variable. This test does not operate by using a mathematical function, but rather by estimating the probability of a value of a test statistic (McDonald 2014). After performing a Fisher's exact test, it is necessary to perform an ad-hoc test if one is making more

than once comparison. For my samples, I will use the Bonferroni correction for multiple tests. In order to determine what p-value is significant using Bonferroni, you divide your normal significant p-value (0.05) by the number of comparisons you are making (McDonald 2014). For my project, I will therefore be using a significant p-value of 0.0042 for my categorical variables.

In order to test whether or not the relationship was significant, a t-test was performed on all variations. A t-test can tell one if there is a significant difference between two populations in terms of one variable. For the purposes of this study, it determines if there is a statistically significant difference in variable X between those individuals with Crohn's disease and those individuals who do not have Crohn's disease. For a t-test,  $T = Z/s$  where Z is  $\sqrt{n}\bar{X}/\sigma$ , where  $\bar{X}$  is the sample mean of the data,  $n$  is the sample size, and  $\sigma$  is the population standard deviation of the data;  $s$  in the one-sample t-test is  $\hat{\sigma}/\sigma$ , where  $\hat{\sigma}$  is the sample standard deviation. The melt temperatures ( $T_M$ ) and cycle threshold ( $C_T$ ) of the samples were analyzed using a non-paired t-test. A nonpaired t-test compares the means of a variable between two different groups to determine if the means are significantly different from one another. The independent samples t-test method compares the mean of a variable for Crohn's disease patients with the mean of that variable in non-Crohn's control samples, as well as ulcerative colitis versus non-ulcerative colitis, and IBD versus non-IBD.

I have provided contingency tables detailing each of the nominal variable tests in Appendix 6. These tables provide the relative risks, the odds ratios, the Fisher's exact p-value, the chi-square p-value, as well as the upper and lower 95% confidence intervals. I have separated them by Crohn's disease vs non-Crohn's tests, UC vs non-UC tests, and IBD vs non-IBD tests. Relative risk is the probability of developing a disease in one group versus the risk of developing a disease if in another group. For example, from my data it would be the risk of developing a



disease (CD, UC, or IBD) if one were female versus if one were male (or using another one of the nominal variables). The Odds Ratio is used to tell the odds that an event or result will happen

compared to the odds of it not occurring (McHugh, 2009). The formula is:

$$\text{Odds ratio} = \frac{PG_1 / (1 - PG_1)}{PG_2 / (1 - PG_2)}$$

“where “PG<sub>1</sub>” is the odds of the event of interest for Group 1, and “PG<sub>2</sub>” represents the odds of the event of interest for group 2” (McHugh, 2009). Odds Ratio is a measure of effect size, meaning that it can tell you the strength of the relationship between two variables. For my data, an OR of 1.00 means that both groups are equally likely to have the disease, higher than 1 means the first is more likely to have the disease, and less than 1 means the first group is less likely to have the disease. Epidemiologists use OR *post hoc* to see if different groups have different outcomes in terms of a particular measure.

## CHAPTER 4: RESULTS

### *PCR RESULTS*

The amplification plots and melt curves for all of the genes after bisulfite conversion are available in Appendix 4. These include the amplification of the *NOD2* promoter region, the *ATG16L1* promoter region, and the *PTPN2* promoter region. The numerical results from the PCR reactions can be found in the statistical analysis sections. The number of samples used for each type of measurement is found in Table 10. A very low melting peak or a late amplification curve will give  $T_m$  values, yet no  $C_t$  values. It is my belief that running 60 cycles rather than 45 cycles would have resulted in a greater number of  $C_t$  values than is shown in the table below.

*Table 10 Number of samples for each measurement broken down by Crohn's disease, ulcerative colitis, and phenotypically normal.*

Measurement	CD samples	UC samples	Control samples	Total
<b>Questionnaire</b>	20	10	41	71
<b>CtNOD2bc</b>	5	4	8	17
<b>TmNOD2bc</b>	5	4	8	17
<b>CtATG16L1bc</b>	2	2	10	14
<b>TmATG16L1bc</b>	2	2	10	14
<b>CtPTPN2bc</b>	4	2	10	16
<b>TmPTPN2bc</b>	4	2	10	16

## STATISTICAL RESULTS

### Means and Standard Deviations

The means and standard deviations for the calculated variables of the *NOD2*, *ATG16L1*, and *PTPN2* genes after bisulfite conversion and post-PCR are given in Tables 11, 12, and 13. Each of these variables are broken down into those samples with Crohn's disease, those samples with ulcerative colitis, and those samples from phenotypically normal individuals. Interval plots of all variables are found in Appendix 5. I ran a test for normality on all of the bisulfite-converted variables using the Anderson-Darling method. All of the variables were normally distributed, except for *ATG16L1*bc Tm.

Table 11 Mean, standard deviation, and number of samples for *NOD2* results. The results are separated into those individuals with Crohn's disease, those with ulcerative colitis, and phenotypically normal controls.

VARIABLE	MEAN	S.D.
<i>NOD2</i> bc Ct Crohn's	41.17	1.56
<i>NOD2</i> bc Ct ulcerative colitis	34.56	1.53
<i>NOD2</i> bc Ct Control	37.85	4.06
<i>NOD2</i> bc Tm Crohn's	71.90	0.07
<i>NOD2</i> bc Tm ulcerative colitis	71.10	0.21
<i>NOD2</i> bc Tm Control	70.89	0.39

Table 12 Mean, standard deviation, and number of samples for *ATG16L1* results. The results are separated into those individuals with Crohn's disease, those with ulcerative colitis, and phenotypically normal controls.

VARIABLE	MEAN	S.D.
<i>ATG16L1</i> bc Ct Crohn's	37.93	2.50
<i>ATG16L1</i> bc Ct ulcerative colitis	40.96	1.58
<i>ATG16L1</i> bc Ct Control	40.45	1.78
<i>ATG16L1</i> bc Tm Crohn's	69.03	0.23
<i>ATG16L1</i> bc Tm ulcerative colitis	69.26	0.12
<i>ATG16L1</i> bc Tm Control	69.21	0.37

Table 13 Mean, standard deviation, and number of samples for PTPN2 results. The results are separated into those individuals with Crohn's disease, those with ulcerative colitis, and phenotypically normal controls.

VARIABLE	MEAN	S.D.
<i>PTPN2</i> bc Ct Crohn's	35.39	0.21
<i>PTPN2</i> bc Ct ulcerative colitis	33.83	1.83
<i>PTPN2</i> bc Ct Control	34.19	2.40
<i>PTPN2</i> bc Tm Crohn's	76.37	0.70
<i>PTPN2</i> bc Tm ulcerative colitis	76.14	0.33
<i>PTPN2</i> bc Tm Control	75.22	1.14

## ***T-TESTS***

### *Crohn's vs non-Crohn's*

The melt temperatures ( $T_M$ ) and cycle threshold ( $C_T$ ) of the samples were analyzed using a 2-sample t-test. When performing a  $t$  test on bisulfite pyrosequencing samples, the 5% reliability limit of the pyrosequencing measurements must be taken into account (Balasa et al. 2010).

For the promotor region of *NOD2* that I chose to amplify, there were no significant differences in means of  $C_t$  values between those individuals with Crohn's disease and those individuals without Crohn's disease, nor were there were significant differences seen between the means of melt temperatures between individuals with Crohn's disease and individuals without the disease (see Table 14). For the promotor region of *ATG16L1*, there were no significant differences in means between those individuals with Crohn's disease and those individuals without Crohn's disease. For the promotor region of *PTPN2*, there were no significant differences in these variables after bisulfite conversion.

Table 14  $T$  test comparing Crohn's disease samples to non-Crohn's disease samples for the *NOD2* gene. Significant  $p$ -values are in red.

<b>Variable</b>	<b><math>T</math></b>	<b><math>p</math></b>	<b>df</b>
<b><i>NOD2bc Ct</i></b>	2.45	0.09	3
<b><i>NOD2bc Tm</i></b>	-2.12	0.08	6
<b><i>ATG16L1bc Ct</i></b>	-2.34	0.052	7
<b><i>ATG16L1bc Tm</i></b>	-1.42	0.18	13
<b><i>PTPN2bc Ct</i></b>	1.93	0.08	11
<b><i>PTPN2bc Tm</i></b>	2.11	0.07	8

### *Ulcerative colitis vs non-UC*

For the promotor region of *NOD2* that I chose to amplify, there were no significant differences between individuals with ulcerative colitis and individuals without ulcerative colitis for any of the three genes (see Table 15).

Table 15 T test comparing ulcerative colitis samples to non-UC samples for the *NOD2* gene. Significant p-values are in red.

Variable	<i>t</i>	<i>p</i>	df
<i>NOD2</i> bc Ct	-2.45	0.09	3
<i>NOD2</i> bc Tm	0.19	0.86	3
<i>ATG16L1</i> bc Ct	1.42	0.23	4
<i>ATG16L1</i> bc Tm	1.21	0.26	9
<i>PTPN2</i> bc Ct	-0.50	0.08	11
<i>PTPN2</i> bc Tm	1.54	0.17	8

*IBD (Crohn's disease + ulcerative colitis) vs non-IBD*

For the promotor region of the three genes being tested, there were no significant differences between individuals with IBD and phenotypically normal individuals (Table 16).

Table 16 T test comparing IBD samples to non-IBD samples for the *NOD2* gene. Significant p-values are in red.

Variable	<i>t</i>	<i>p</i>	df
<i>NOD2</i> bc Ct	0.01	0.99	5
<i>NOD2</i> bc Tm	2.25	0.09	4
<i>ATG16L1</i> bc Ct	-1.41	0.18	14
<i>ATG16L1</i> bc Tm	-0.65	0.53	11
<i>PTPN2</i> bc Ct	0.76	0.46	13
<i>PTPN2</i> bc Tm	2.49	0.07	13

### ***CORRELATIONS AND FISHER'S EXACT TESTS***

A correlation was run between each variable and either Crohn's disease, ulcerative colitis, or phenotypically normal controls. The p-value given was determined after a Bonferroni test (when there were equal variances) or a Tamhane test (where there were not equal variances) to correct for multiple comparisons.

#### *Crohn's disease Correlations:*

For the *NOD2*, *ATG16L1*, and *PTPN2* bisulfite-converted melt temperatures and Ct values, there were no significant correlations with Crohn's disease after Bonferroni multiple test correction.

*Table 17 Correlation between each of the NOD2 variables and Crohn's disease. Significant p-values are in red.*

<b>Correlated with Crohn's disease</b>	<b>Correlation</b>	<b>p value</b>
<i>NOD2bc Ct</i>	0.36	0.31
<i>NOD2bc Tm</i>	-0.72	0.15
<i>ATG16L1bc Ct</i>	0.56	0.19
<i>ATG16L1bc Tm</i>	0.32	0.24
<i>PTPN2bc Ct</i>	-0.28	0.38
<i>PTPN2bc Tm</i>	-0.41	0.88
<b>Age</b>	0.23	0.15

Fisher's exact test was performed on Crohn's disease and the variables from the questionnaire that had two responses (i.e. male/female or yes/no). There were no significant differences between Crohn's disease and any of the variables. All of the p-values from this test may be seen in Table 18.

*Table 18 Fisher's exact test comparing CD to variables. Significant p-values <0.0042 (Bonferroni correction) are in red.*

<b>Fisher's exact test for CD</b>	<b>p value</b>
Gender	0.78
Environment	0.41

Smoker	0.99
Alcohol use	0.99
Allergies	0.03
Familial Crohn's disease	0.01
Heart attack	0.32
Cancer	0.27
Celiac disease	0.57
High cholesterol	0.05
Asthma	0.72
Rheumatoid arthritis	0.40

#### ***ULCERATIVE COLITIS CORRELATIONS:***

For the *NOD2*, *ATG16L1*, and *PTPN2* promoter region, there were no significant correlations between ulcerative colitis and any of the bisulfite converted variables (Table 19).

*Table 19 Correlations between each of the NOD2 variables and ulcerative colitis. Significant p-values are in red.*

<b>Correlated with ulcerative colitis</b>	<b>Correlation</b>	<b>p value</b>
<i>NOD2</i> bc Ct	0.36	0.30
<i>NOD2</i> bc Tm	0.03	0.99
<i>ATG16L1</i> bc Ct	-0.27	0.97
<i>ATG16L1</i> bc Tm	-0.17	0.98
<i>PTPN2</i> bc Ct	0.12	0.99
<i>PTPN2</i> bc Tm	-0.19	0.79
Age	-0.06	0.83

Fisher's exact test was performed on ulcerative colitis and the variables from the questionnaire that had two responses (i.e. male/female or yes/no). There were no significant differences between ulcerative colitis and any of the variables. All of the p-values from this test may be seen in Table 20.

*Table 20 Fisher's exact test comparing UC to variables. Significant p-values <0.0042 are in red.*

<b>Fisher's exact test for UC</b>	<b>p value</b>
Gender	0.49



Environment	0.73
Smoker	0.45
Alcohol use	0.73
Allergies	0.17
Familial Crohn's disease	0.58
Heart attack	0.99
Cancer	0.99
Celiac disease	0.46
High cholesterol	0.43
Asthma	0.99
Rheumatoid arthritis	0.69

#### ***IBD CORRELATIONS:***

There were no significant correlations between the *NOD2*, *ATG16L1* and *PTPN2* variables and IBD (see Table 21).

*Table 21 Correlations between each of the NOD2 variables and IBD. Significant p-values are in red.*

<b>Correlated with IBD</b>	<b>Correlation</b>	<b>p value</b>
<b><i>NOD2</i>bc Ct</b>	0.01	0.99
<b><i>NOD2</i>bc Tm</b>	0.58	0.13
<b><i>ATG16L1</i>bc Ct</b>	0.36	0.32
<b><i>ATG16L1</i>bc Tm</b>	0.17	0.88
<b><i>PTPN2</i>bc Ct</b>	-0.17	0.72
<b><i>PTPN2</i>bc Tm</b>	-0.49	0.24
<b>Age</b>	0.25	0.13

Fisher's exact test was performed on ulcerative colitis and the variables from the questionnaire that had two responses (i.e. male/female or yes/no). The only significance found was between IBD and high cholesterol. Individuals with IBD had significantly less diagnoses of high cholesterol than phenotypically normal individuals. All of the p-values from this test may be seen in Table 22.

*Table 22 Fisher's exact test comparing IBD to variables. Significant p-values <0.0042 are in red.*

<b>Fisher's exact test for IBD</b>	<b>p value</b>
Gender	0.81
Environment	0.22
Smoker	0.44
Alcohol use	0.99
Allergies	0.81
Familial Crohn's disease	0.12
Heart attack	0.07
Cancer	0.10
Celiac disease	0.99
High cholesterol	<b>0.0041</b>
Asthma	0.34
Rheumatoid arthritis	0.78

## CHAPTER 5: DISCUSSION AND CONCLUSIONS

The statistical tests used in this project showed few significant differences between Crohn's disease, ulcerative colitis, and phenotypically normal samples after correcting for multiple testing. There were no obvious differences between Crohn's disease, ulcerative colitis, and phenotypically normal individuals that could be used as a method for differentiating between the groups. The only significant differences found were correlations between IBD and high cholesterol (a negative correlation).

Methylation of a gene causes the Ct value to be lower and the Tm to be higher than is present in a non-methylated gene. Methylation of a promoter region can cause either increased or decreased expression of a gene. When looking at the 95% confidence intervals (Appendix 5) of means, Crohn's disease had a melt temperature mean that was different than the confidence interval of non-Crohn's disease individuals (including those with ulcerative colitis) for the *NOD2* Tm. It should be noted, however that there were only two samples of Crohn's disease used in the analysis of this variable, thus not giving it enough statistical power to definitively say there are differences. Thus, the null hypothesis that CD and UC samples are within the 95% confidence interval of the control samples cannot be discarded for the three genes in question.

The 2-sample *t*-test table showed which variables have means that are significantly different. None of the variables were significantly different in Crohn's disease versus non-Crohn's disease. None of the variables were significant for ulcerative colitis or IBD, either.

None of the variables (either measured or from the questionnaire) were significantly correlated with ulcerative colitis or IBD. The only variables that were correlated with Crohn's disease were: familial history of Crohn's disease and high cholesterol.

This study found that people with Crohn's disease were significantly more likely to have less diagnoses of high cholesterol than individuals with ulcerative colitis and to have a family member with Crohn's disease as opposed to individuals without the disease. The mean age of diagnosis for Crohn's disease is lower than the mean age of diagnosis for ulcerative colitis. It should be noted that there are two peaks of incidence for ulcerative colitis – one at age 15-30 and another at age 50-65. Thus, while the vast majority of individuals with Crohn's disease have been diagnosed by the age 35, many individuals with ulcerative colitis do not begin showing symptoms until a much later age. Due to the fact that there was a significant difference in age between Crohn's disease patients and ulcerative colitis patients, it is not surprising that Crohn's disease was also found to be associated with less high cholesterol, which tends to be diagnosed more in older individuals. Once again, it is important to stress that correlation does not equal causation.

The *PTPN2* gene produces a protein that is essential for distinguishing normal intestinal flora from pathogenic bacteria. Abnormalities in this gene can lead to a decrease in bacterial sensing, thus allowing an increase in pathogenic bacteria in the intestines. The *NOD2* gene limits pro-inflammatory effects in the intestines and regulates autophagy. Abnormalities in the gene can result in enhanced T-cell proliferation, increased inflammation, and an increase in bacterial invasion that can lead to even greater levels of inflammation in the intestines. The *ATG16L1* gene is expressed in intestinal epithelial cells and involved in the intracellular autophagy complex. Abnormalities in this gene can impair the autophagy ability of cells, and diminish bacterial clearance. These three genes essentially work in concert by identifying pathogenic bacteria, forming a response, and ultimately clearing out the bacteria. Abnormalities in any of these genes (as well as the multitude of other genes associated with Crohn's disease and

ulcerative colitis), whether mutations or epigenetic effects, can thus impair the system, which is what is seen in Crohn's disease and ulcerative colitis. Based on the research in this dissertation, it was not found that regions of the *NOD2*, *ATG16L1* and *PTPN2* genes are differently methylated between IBD and non-IBD individuals.

It was my hope that by looking at the promoter regions of genes known to be associated with Crohn's disease that I would be able to develop a non-invasive method for differentiating Crohn's disease from ulcerative colitis from phenotypically normal individuals. Alas, this was not the case. One of the most recently published research papers looked at methylation genome-wide. They found that the majority of DNA methylation in individuals with Crohn's disease was actually found within introns and intergenic regions, with only a very low percentage occurring in the promoter regions or exons of genes (Sadler et al. 2017). This may explain the negative results I received during my own research. Based on the limited number of samples that I used for this project, there does not appear to be differences in methylation status between the three groups I was examining. Future research that I conduct will likely either look at places outside the promoter regions, or else the promoter regions of the few genes that Sadler et al found to be differentially methylated. It is possible that some of those regions may be used to devise a more accurate, less invasive diagnostic tool for IBD.

This dissertation has **not** shown that there are significant genetic and epigenetic differences between Crohn's disease, ulcerative colitis, and phenotypically normal individuals. It is possible that there are differences, but the small sample sizes used here were unable to detect them. This project had small sample sizes. In the future, it is my intention to continue this research with larger sample as well as additional genes, additional environmental factors, expanded family medical histories, and information on the health and nutrition of the

individual's mother during pregnancy. While my results did not show any obvious epigenetic differences between the groups for any of the genes being tested, it is still my belief that there may be differential methylation somewhere in the genome that can differentiate Crohn's disease, ulcerative colitis, and phenotypically normal individuals. It is important that an individual be correctly diagnosed as the treatment plans for these two types of irritable bowel disease differ in type and efficacy.

## REFERENCES

- Abbas AK, Lichtman AH, and Pillai S. 2014. Basic Immunology: Functions and Disorders of the Immune System, 4th Edition. Philadelphia: Elsevier Saunders.
- Adam M, Polifka J, and Friedman J. 2011. Evolving Knowledge of the Teratogenicity of Medications in Human Pregnancy. *American Journal of Medical Genetics C Semin Med Genet* 157C(3):175-182.
- Adams A, Kennedy N, Hansen R, Ventham N, O'Leary K, Drummond H, Noble C, El-Omar E, Russell R, Wilson D et al. . 2014a. Two-stage Genome-wide Methylation Profiling in Childhood-onset Crohn's Disease Implicates Epigenetic Alterations at the VMP1/MIR21 and HLA Loci. *Inflammatory Bowel Disease*.
- Adams AT, Kennedy NA, Hansen R, Ventham NT, O'Leary KR, Drummond HE, Noble CL, El-Omar E, Russell RK, Wilson DC et al. . 2014b. Two-stage Genome-wide Methylation Profiling in Childhood-onset Crohn's Disease Implicates Epigenetic Alterations at the VMP1/MIR21 and HLA Loci. *Inflammatory Bowel Disease* 20(10):1784-1793.
- Adams DD, and Adams CD. 2013. Autoimmune Disease Pathogenesis, Genetics, Immunotherapy, Prophylaxis, and Principles for Organ Transplantation. Dordrecht, Heidelberg, New York, London: Springer.
- Alexander LL. 2014. Autoimmune Diseases. CME.
- American Autoimmune Related Diseases Association I. 2014. Autoimmune Statistics.
- Andraos C, Koorsen G, Knight JC, and Bornman L. 2011. Vitamin D Receptor Gene Methylation is Associated with Ethnicity, Tuberculosis, and *TaqI* Polymorphism. *Humman Immunology* 72:262-268.
- Annunziata ML, Caviglia R, Micera A, Papparella LG, and Cicala M. 2011. Expression of GH/IGF-1 Axis Compounds in Moderate-Severe Crohn's Disease: Role of Mucosal Inflammation and Anti TNF- $\alpha$  Treatment. *American Gastroenterological Association Meeting*. Chicago, IL. p Su1223.
- Anway MD, Cupp AS, Uzumcu M, and Skinner MK. 2005. Epigenetic Transgenerational Actions of Endocrine Disruptors and Male Fertility. *Science* 308:1466-1470.
- Arai T, Kakuta Y, Kinouchi Y, Negoro K, Endo K, Shiga H, Takahashi S, and Shimosegawa T. 2011. Increased Expression of NKX2.3 mRNA Transcribed From Risk Haplotype for Inflammatory Bowel Disease in the Involved Colonic Mucosa. *American Gastroenterological Association Meeting*. Chicago, IL. p Sa1272.
- Arijs I, Van Lommel L, De Hertogh G, Machiels K, Lemaire K, Van Steen K, Van Assche GA, Vermeire S, Schult FC, and Rutgeerts PJ. 2011. Mucosal Gene Expression Profiling Differentiates Early From Late Ileal Crohn's Disease. *American Gastroenterological Association Meeting*. Chicago, IL. p Sa1273.
- Arnold C, F. SP, and Prohaska SJ. 2013. Chromatin Computation: epigenetic inheritance as a pattern reconstruction problem. *Journal of Theoretical Biology* 336:61-74.
- Bach J. 2002. The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases. *New England Journal of Medicine* 347(12):911-920.
- Backdahl L. 2010. Epigenetic Regulation of the Inflammatory Response. *5th International Meeting on Inflammatory Bowel Diseases*. Capri. p E15.
- Backdahl L, Bushell A, and Beck S. 2009. Inflammatory Signalling as Mediator of Epigenetic Modulation in Tissue-specific Chronic Inflammation. *The International Journal of Biochemistry & Cell Biology* 41:176-184.

- Bai AH, Wu WK, Xu L, Wong SH, Go MY, Chan AW, Harbord M, Zhang S, Chen M, Wu JC et al. . 2017. Dysregulated Lysine Acetyltransferase 2B Promotes Inflammatory Bowel Disease Pathogenesis Through Transcriptional Repression of Interleukin-10. *Journal of Crohn's and Colitis* 10(6):726-734.
- Balasa A, Gathungu G, Kisfali P, Smith EOb, Cho JH, Melagh B, and Kellermayer R. 2010. Assessment of DNA Methylation at the Interferon Regulatory Factor 5 (IRF5) Promoter Region in Inflammatory Bowel Diseases. *International Journal of Colorectal Disease* 0:000-000.
- Barnett M, Bermingham E, McNabb W, Bassett S, Armstrong K, Rounce J, and Roy N. 2010. Investigating Micronutrients and Epigenetic Mechanisms in Relation to Inflammatory Bowel Disease. *Mutation Research* 690:71-80.
- Barres R, and Zierath JR. 2011. DNA Methylation in Metabolic Disorders. *American Journal of Clinical Nutrition* 93:897S-900S.
- Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM et al. . 2008. Genome-wide Association Defines More Than 30 Distinct Susceptibility Loci for Crohn's Disease. *Nature Genetics* 40(8):955-962.
- Barrett M, and Chandra SB. 2011. A Review of Major Crohn's Disease Susceptibility Genes and Their Role in Disease Pathogenesis. *Genes & Genomics* 33:317-325.
- Barrios RJ, Kheradmand F, Batts L, and Corry DB. 2006. Asthma: Pathology and Pathophysiology. *Arch Pathol Lab Med* 130:447-452.
- Batra A, Kredel LI, Freise I, Zeitz M, and Siegmund B. 2011. Regulation of Macrophage Phenotype in Crohn's Disease Hypertrophied Adipose Tissue. *American Gastroenterological Association Meeting*. Chicago, IL. p Su1745.
- Bayarsaihan D. 2011. Epigenetic Mechanisms in Inflammation. *Journal of Dental Research* 90(1):9-17.
- Behr MA. 2010. The Path to Crohn's Disease: Is Mucosal Pathology a Secondary Event? *Inflammatory Bowel Disease* 000:000-000.
- Binder V, and Orholm V. 1996. Familial Occurrence and Inheritance Studies in Inflammatory Bowel Disease. *Netherlands Journal of Medicine* 48(2):53-56.
- Bird A. 2007. Perceptions of Epigenetics. *Nature* 447:396-398.
- Bird A, Tate P, Nan X, Campoy J, Meehan R, Cross S, Tweedie S, Charlton J, and Macleod D. 1995. Studies of DNA Methylation in Animals. *Journal Cell Science Suppl* 19:37-39.
- Boder M, Kelly EJ, and Ho RJ. 2005. Characterization of the Human MDR1 Gene. *Journal of the American Association of Pharmaceutical Scientists* 7:E1-E5.
- Bonen DK, and Cho JH. 2003. The Genetics of Inflammatory Bowel Disease. *Gastroenterology* 126:625-627.
- Bossdorf O, Richards CL, and Pigliucci M. 2008. Epigenetics for Ecologists. *Ecological Letters* 11:106-115.
- Brasset E, and Chambeyron S. 2013. Epigenetics and Transgenerational Inheritance. *Genome Biology* 14:306-308.
- Bridger S, Lee JCW, Bjarnason I, Lennard Jones JE, and Macpherson AJ. 2002. In Siblings with Similar Genetic Susceptibility for Inflammatory Bowel Disease, Smokers Tend to Develop Crohn's Disease and Non-smokers Develop Ulcerative Colitis. *Gut* 51:21-25.
- Brix TH, Hansen PS, Bennedbaek FN, Bonnema SJ, Kyvik KO, Orstavik KH, and Hegedus L. 2009. X Chromosome Inactivation Pattern is not Associated With Interindividual



- Variations in Thyroid Volume: A Study of Euthyroid Danish Female Twins. *Twin Research and Human Genetics* 12(5):502-506.
- Brooks WH, Le Dantec C, Pers J-O, Youinou P, and Renaudineau Y. 2010. Epigenetics and Autoimmunity. *Journal of Autoimmunity* 34:J207-J219.
- Burakoff R, Hande S, Ma J, Banks PA, Friedman S, Makrauer F, and Liew C-C. 2010. Differential Regulation of Peripheral Leukocyte Genes in Patients With Active Crohn's Disease and Crohn's Disease in Remission. *Journal of Clinical Gastroenterology* 44:120-126.
- Burdge GC, and Lillycrop KA. 2010. Nutrition, Epigenetics, and Developmental Plasticity: Implications for Understanding Human Disease. *Annual Review Nutrition* 30:315-339.
- Callinan PA, and Feinberg AP. 2006. The Emerging Science of Epigenomics. *Human Molecular Genetics* 15(1):95-101.
- Champagne FA, and Curley JP. 2009. Epigenetic Mechanisms Mediating the Long-Term Effects of Maternal Care on Development. *Neuroscience and Biobehavioral Reviews* 33:593-600.
- Cheon JH. 2013. Genetics of Inflammatory Bowel Diseases: a comparison between Western and Eastern perspectives. *Journal of Gastroenterology and Hepatology* 28:220-226.
- Cho JH. 2001. Update on the Genetics of Inflammatory Bowel Disease. *Current Gastroenterology* 3:458-463.
- Cho JH, and Brant SR. 2011. Recent Insights into the Genetics of Inflammatory Bowel Disease. *Gastroenterology* 140:1704-1712.
- Choi CQ. 2011. Having a Baby Makes Mom's Body Turn on Itself. *PLoS ONE*.
- Choi JK, and Kim SC. 2007. Environmental Effects on Gene Expression Phenotype Have Regional Biases in the Human Genome. *Genetics* 175:1607-1613.
- Clark H, Pidasheva S, Trifari S, Hackney J, Ma Y, Smith A, Sohn S, Spits H, Little R, Behrens T et al. . 2011. Human IL23R R381Q Genetic Variant Implicated in Crohn's Disease Pathogenesis is Hypomorphic Resulting in Reduced Function of the Receptor. *American Gastroenterological Association Meeting*. Chicago, IL. p Sa1288.
- Cleynen I, Artieda M, Szczypiorska M, Verspaget HW, Lakatos PL, Seibold F, Ahmad T, Weersma RK, Muller S, Tordai A et al. . 2011. The Protease Genes *Cyld* and *USP40* are Associated with Crohn's Disease: results from a European Consortium. *American Gastroenterological Association Meeting*. Chicago, IL. p Sa1268.
- Cleynen I, John JMM, Henckaerts L, Van Moerkercke W, Rutgeerts P, Van Steen K, and Vermeire S. 2010. Molecular Reclassification of Crohn's Disease by Cluster Analysis of Genetic Variants. *PLoS One* 5(9):e12952-e12959.
- Colman RJ, and Rubin DT. 2014. Fecal Microbiota Transplantation as Therapy for Inflammatory Bowel Disease: a systemic review and meta-analysis. *Journal of Crohn's and Colitis* 8(12):1569-1581.
- Connor CM, and Akbarian S. 2008. DNA Methylation Changes in Schizophrenia and Bipolar Disorder. *Epigenetics* 3(2):55-58.
- Consortium MS. 1999. Complete Sequence and Gene Map of a Human Major Histocompatibility Complex. *Nature*. 401 (6756): 921-923.
- Cooke J, Zhang H, Greger L, Silva A-L, Massey D, Dawson C, Metz A, Ibrahim A, and Parkes M. 2012. Mucosal Genome-wide Methylation Changes in Inflammatory Bowel Disease. *Inflammatory Bowel Disease* 18(11):2128-2137.
- Cooney CA. 2007. Epigenetics - DNA-Based Mirror of Our Environment? *Disease Markers* 23:121-137.

- Cooper GS, WMiller FW, and Pandey JP. 1999. The Role of Genetic Factors in Autoimmune Disease: implications for environmental research. *Environmental Health Perspectives* 107(5):693-700.
- Costello CM, Mah N, Hasler R, Rosenstiel P, Waetzig GH, Hahn A, Lu T, Gurbuz Y, Nikolaus S, Albrecht et al. . 2005. Dissection of the Inflammatory Bowel Disease Transcriptome Using Genome-Wide cDNA Microarrays. *PLoS Medicine* 2:771-787.
- Curley JP, Mashoodh R, and Champagne FA. 2011. Epigenetics and the Origins of Paternal Effects. *Hormones and Behavior* 59:306-314.
- Dabritz J, and Menheniott TR. 2014. Linking Immunity, Epigenetics, and Cancer in Inflammatory Bowel Disease. *Inflammatory Bowel Disease* 0:1-17.
- Dave M, Gorospe EC, and Luter J. 2011. Faecalibacterium Prausnitzii in Crohn's Disease: Hope or Hype? A Systemic Review. *American Gastroenterological Association Meeting*. Chicago, IL. p Sa1793.
- de Hertogh G, and Goeboes K. 2004. Crohn's Disease and Infections: A Complex Relationship. *Journal of Medical Genetics* 6(3):14-17.
- Dessein R, Chamaillard M, and Danese S. 2008. Innate Immunity in Crohn's Disease: the reverse side of the medal. *Journal of Clinical Gastroenterology* 42:S144-S147.
- Dhôte R, Dhôte-Burger P, Thevenot T, Devaux J-Y, Beades E, Richard B, and Christoforov B. 2000. Polymorphonuclear Leukocyte-elastase in Crohn's Disease. *Journal of Clinical Gastroenterology* 31(2):152-155.
- Di Sabatino A, Rovedatti L, Vidali F, MacDonald TT, and Corazza GR. 2013. Recent Advances in Understanding Crohn's Disease. *Internal and Emergency Medicine* 8:101-113.
- Dimitroulia E, Pitiriga VC, Piperaki E-T, Spanakis NE, and Tsakris A. 2013. Inflammatory Bowel Disease Exacerbation Associated with Epstein-Barr Virus Infection. *Disease Colon Rectum* 56:322-327.
- Dinu I, Mahasirimongkol S, Liu Q, Yanai H, Eldin NS, Kreiter E, Wu X, Jabbari S, Tokunaga K, and Yasui Y. 2012. SNP-SNP Interactions Discovered by Logic Regression Explain Crohn's Disease Genetics. *PLoS One* 7(10):e43035.
- Drouet M, Vignal C, Singer E, Dubreuil L, Desreumaux P, and Neut C. 2011. Antibiotics Induced Commensal Flora Disruption Favours *Escherichia coli* AIEC (LF82) Colonization in Wild Type and NOD2 Knock-out Mice. *American Gastroenterological Association Meeting*. Chicago, IL. p Sa1773.
- Dubinsky M, kugathan S, Kwon S, Haritunians T, Wrobel IT, Wahbeh G, Quiros A, Bahar RJ, Farrior S, Teleten N et al. . 2011. A Combination of Genetic, Clinical and Immune Markers Predict the Need for Surgery in Crohn's Disease. *American Gastroenterological Association Meeting*. Chicago, IL. p 934.
- Dubois P, Degraeve E, and Vandeplas O. 1998. Asthma and Airway Hyperresponsiveness among Belgian Conscripts, 1978-91. *Thorax* 53:101-105.
- Dunkelberger JR, and Song W-C. 2010. Complement and its Role in Innate and Adaptive Immune Responses. *Cell Research* 20:34-50.
- Elliott T, Prescott NJ, Hudspith BN, Rayment NB, Hermon-Taylor J, Brostoff J, Mathew CG, and Sanderson JD. 2011. Autophagy Gene Polymorphisms Influence the Interaction of *E. coli* and Macrophages in Crohn's Disease. *American Gastroenterological Association Meeting*. Chicago, IL. p 112.
- Fabio RR, Concetta RM, Giuseppe C, Sara R, and Ambrogio O. 2011a. Contribution to Crohn's Disease Risk in Sicily. *Inflammatory Bowel Disease* 17:1635-1636.

- Fabio RR, Concetta RM, Giuseppe C, Sara R, Ambrogio O, Aurelio M, and Mario C. 2011b. ATG16L1 Contribution to Crohn's Disease Risk in Sicily. *Inflamm Bowel Dis* 17(7):1635-1636.
- Farmer M, Hunt L, Eichenberger MR, Petras RE, Janosky JE, and Galandiuk S. 2001. Association of Susceptibility Locus for Inflammatory Bowel Disease on Chromosome 16 with Both Ulcerative Colitis and Crohn's Disease. *Digestive Diseases and Sciences* 46:632-636.
- Feagan B, Vreeland M, Larson L, and Bala M. 2000. Annual Cost of Care for Crohn's Disease: a payor perspective. *American Journal of Gastroenterology* 95(8):1955-1960.
- Feinberg AP. 2007. Phenotypic Plasticity and the Epigenetics of Human Disease. *Nature* 447:433-442.
- Feinberg AP, and Irizarry RA. 2009. Stochastic Epigenetic Variation as a Driving Force of Development, Evolutionary Adaptation, and Disease. *PNAS*(10):1-8.
- Fell JM, Paintin M, Arnaud-Battandier F, Beattie R, Hollis A, Kitching P, Donnet-Hughes A, MacDonald TT, and Walker-Smith JA. 2000. Mucosal Healing and a Fall in Mucosal Pro-Inflammatory Cytokine mRNA Induced by a Specific Oral Polymeric Diet in Paediatric Crohn's Disease. *Alimentary Pharmacology Therapy* 14:281-289.
- Festen EAM, Goyette P, Green T, Boucher G, Beauchamp C, Trynka G, Dubois PC, Legace C, Stokkers PCF, Hommes DW et al. . 2011. A Meta-Analysis of Genome-Wide Association Scans Identifies IL18RAP, PTPN2, TAGAP, and PUS10 As Shared Risk Loci for Crohn's Disease and Celiac Disease. *PLoS Genetics* 7(1):e1001283-e1001288.
- Finney OC, Nwakanma D, Conway DJ, Walther M, and Riley EM. 2009. Homeostatic Regulation of T Effector to Treg Ratios in an Area of Seasonal Malaria Transmission. *European Journal of Immunology* 39(5):1288-1300.
- Flintoft L. 2005. Identical Twins: Epigenetics Makes the Difference. *Nature Reviews* 6:667-668.
- Floor M, Binion DG, Nelson VM, Manley S, Wellner M, Sadeghi S, Behmaram B, Sewell C, Otterson MF, Kucharz T et al. . 2008. Role of MutS Homolog 2 (MSH2) in Intestinal Myofibroblast Proliferation during Crohn's Disease Stricture Formation. *American Journal of Gastrointestinal and Liver Physiology* 295:G581-G590.
- Fofanova TY, Petrosino JF, and Kellermayer R. 2016. Microbiome-Epigenome Interactions and the Environmental Origins of Inflammatory Bowel Diseases. *Journal of Pediatric Gastroenterology Nutrition* 62(2):208-219.
- Forabosco P, Collins A, Latiano A, Annese V, Clementi M, Andriulli A, Fortina P, Devoto M, and Morton NE. 2000. Combined Segregation and Linkage Analysis of Inflammatory Bowel Disease in the IBD1 Region Using Severity to Characterise Crohn's Disease and Ulcerative Colitis. *European Journal of Human Genetics* 8:846-852.
- Franke A, Balschun T, Karlsen TH, Hedderich J, May S, Lu T, Schuldt D, Nikolaus S, Rosenstiel P, Krawczak M et al. . 2008. Replication of Signals from Recent Studies of Crohn's Disease Identifies Previously Unknown Disease Loci for Ulcerative Colitis. *Nature Genetics* 40(6):713-715.
- Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, P.L. M, and Paul CL. 1992. A Genomic Sequencing Protocol that Yields a Positive Display of 5-methylcytosine Residues in Individual DNA Strands. *Proceedings of the National Academy of Science* 89(5):1827-1831.
- Gabbani T, Deiana S, Marocchi M, and Annese V. 2017. Genetic Risk Variants as Therapeutic Targets for Crohn's Disease. *Pharmacology & Pharmacy* 21(4):381-390.

- Gavin DP, Kartan S, Chase K, Jayaraman S, and Sharma RP. 2009. Histone Deacetylase Inhibitors and Candidate Gene Expression: an in vivo and in vitro approach to studying chromatin remodeling in a clinical population. *Journal of Psychiatric Research* 43:870-876.
- Gevers D, Kugathasan S, and Denson L. 2014. Gut Microbiome in New-onset Crohn's Disease. *Cell Host and Microbe* 15:332-392.
- Giallourakis C, Stoll M, Miller K, Hampe J, Lander ES, Daly MJ, Schreiber S, and Rioux JD. 2003. IBD5 is a General Risk Factor for Inflammatory Bowel Disease: Replication of Association with Crohn's Disease and Identification of a Novel Association with Ulcerative Colitis. *American Journal of Human Genetics* 73:205-211.
- Glas J, Seiderer J, Wetzke M, Konrad A, Torok HP, Schmechel S, Tonenchi L, Grassl C, Dambacher J, Pfennig S et al. . 2007. rs1004819 is the Main Disease-Associated Variant in German Crohn's Disease Patients: Combined Analysis of IL23R, CARD15, and OCTN1/2 Variants. *PLoS ONE* 2(9):e819.
- Gonsky R, Deem R, Landers C, Haritunians T, Yang S, and Targan S. 2014. IFNG rs1861494 Polymorphism is Associated with IBD Disease Severity and Functional Changes in Both IFNG Methylation and Protein Secretion. *Inflammatory Bowel Disease*.
- Greco L, Romino R, Coto I, Cosmo ND, Percopo S, Maglio F, and Stazi MA. 2002. The First Large Population Based Twin Study of Coeliac Disease. *Gut* 50(5):624-628.
- Grynberg M, Graesslin O, Teyssedre J, Quereux C, Gaillard D, and Carre-Pigeon F. 2007. Prenatal Diagnosis in Monozygotic Twins with Down Syndrome who had Different Phenotypes. *Prenatal Diagnosis* 27:552-554.
- Gupta R, Nagarajan A, and Wajapeyee N. 2010. Advances in Genome-wide DNA Methylation Analysis. *BioTechniques* 49(iii-xi):iii-xi.
- Gynecologists TACoOa. 2013.
- Haig D. 2006. Weismann Rules! OK? Epigenetics and the Lamarckian Temptation. *Biology and Philosophy* 22:415-428.
- Halfvarson J. 2011. Genetics in Twins with Crohn's Disease: less pronounced than previously believed? *Inflammatory Bowel Disease* 17(1):6-12.
- Hamdan-Khalil R, Gala J-L, Allorge D, Lo-Guidice J-M, Horsmans Y, Houdret N, and Broly F. 2005. Identification and Functional Analysis of Two Rare Allelic Variants of the Thiopurine S-methyltransferase Gene, TPMT\*16 and TPMT\*19. *Biochemical Pharmacology* 69:525-529.
- Harris RA, Nagy-Szakal D, Pedersen N, Opekun A, Bronsky J, Munkholm P, Jespersgaard C, Andersen P, Melegh B, Ferry G et al. . 2012. Genome-wide Peripheral Blood Leukocyte DNA Methylation Microarrays Identified a Single Association with Inflammatory Bowel Diseases. *Inflammatory Bowel Disease* 18(12):2334-2342.
- Hasler R, Feng Z, and Backdahl L. 2012. A Functional Methylome Map of Ulcerative Colitis. *Genome Research* 22(2130-2137).
- Hayakawa K, Kato K, Onoi M, Yang-Ping C, Kanamori M, Doi S, Hayashi C, Kikuchi H, Nishihara R, and Kadota K. 2006. The Osaka University Aged Twin Registry: Epigenetics and Identical Twins Discordant for Aging-Dependent Diseases. *Twin Research and Human Genetics* 9(6):808-810.
- Heijmans B, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, and Lumey LH. 2008a. Persistent Epigenetic Differences Associated with Prenatal Exposure to Famine in Humans. *PNAS* 105(44):17046-17049.

- Heijmans BT, Kremer D, Tobi EW, Boomsma DI, and Slagboom PE. 2007. Heritable Rather than Age-related Environmental and Stochastic Factors Dominate Variation in DNA Methylation of the Human IGF2/H19 Locus. *Human Molecular Genetics* 16(5):547-554.
- Heijmans BT, Tobi EW, Stein AD, Putter H, and Blauw GJ. 2008b. Persistent Epigenetic Differences Associated to Famine in Humans. *PNAS* 105:17046-17049.
- Heresbach D, Alizadeh M, Dabadie A, Le Berre N, Colombel JF, Yaouanq J, Bretagne JF, and Semana G. 1997. Significance of Interleukin-1 beta and Interleukin-1 Receptor Antagonist Genetic Polymorphism in Inflammatory Bowel Diseases. *American Journal of Gastroenterology* 92:1164-1169.
- Hermon-Taylor J. 2009. *Mycobacterium avium* subspecies paratuberculosis, Crohn's Disease, and the Doomsday Scenario. *Gut Pathogens* 1:1-15.
- Hermus JPS, van Rhijn LW, and van Ooij A. 2007. Non-Genetic Expression of Adolescent Idiopathic Scoliosis: a case report and review of the literature. *European Spine Journal* 16(3):338-341.
- Herrlinger KR, Fellermann K, Fischer C, Kreisel W, Deibert P, Schoelmerich J, Fleig WE, Ruhl A, Reinshagen M, Greinwald R et al. . 2004. Thioguanine-nucleotides Do Not Predict Efficacy of Thioguanine in Crohn's Disease. *Alimentary Pharmacology Therapy* 19:1269-1276.
- Holoch D, and Moazed D. 2015. RNA-Mediated Epigenetic Regulation of Gene Expression. *Nature Reviews Genetics* 16:71-85.
- Honzawa Y, Nakase H, Matsumura K, Yamamoto S, Uza N, Matsuura M, and Chiba T. 2011. IL-17 Promotes HSP47 Expression and Intestinal Fibrosis in Crohn's Disease. *American Gastroenterological Association Meeting*. Chicago, IL. p Su1772.
- Horst SN, Schwartz DA, Beaulieu DB, Duley C, Wagnon JH, Tossberg JT, Henderson MA, Crooke P, and Aune TM. 2011. A Pilot Study Exploring the Utility of Gene Expression Profiling to Differentiate Crohn's Disease from Ulcerative Colitis and IBD from Irritable Bowel Syndrome in Patients with Chronic Diarrhea. *American Gastroenterological Association Meeting*. Chicago, IL. p Mo1978.
- Hu J, and Peter I. 2013. Evidence of Expression Variation and Allelic Imbalance in Crohn's Disease Susceptibility Genes NOD2 and ATG16L1 in Human Dendritic Cells. *Gene* 527:496-502.
- Hubbard V, and Cadwell K. 2011. Viruses, Autophagy Genes, and Crohn's Disease. *Viruses* 3:1281-1311.
- Huff CD, Witherspoon DJ, Zhang Y, Gatenbee C, Denson LA, Kugathasan S, Hakonarson H, Whiting A, Davis CT, Wu W et al. . 2002. Crohn's Disease and Genetic Hitchhiking at *IBD5*. *Molecular Biology Evolution* 29(1):101-111.
- Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M et al. . 2001. Association of NOD2 Leucine-Rich Repeat Variants with Susceptibility to Crohn's Disease. *Nature* 411:599-603.
- Hutchinson JN, Raj T, Fagerness J, Stahl E, Vilorio FT, Gimelbrant A, Seddon J, Daly M, Chess A, and Plenge R. 2014. Allele-Specific Methylation Occurs at Genetic Variants Associated with Complex Disease. *PLoS ONE* e98464.
- Innan H, and Kim Y. 2004. Pattern of Polymorphism after Strong Artificial Selection in a Domestication Event. *Proceedings of the National Academy of Sciences* 101:6473-6481.

- Intine RV, Olsen AS, and Sarras MP. 2013. A Zebrafish Model of Diabetes Mellitus and Metabolic Memory. *Journal of Visualized Experiments* 72:1-7.
- Invernizzi P, and Gershwin ME. 2009. The Genetics of Human Autoimmune Disease. *Journal of Autoimmunity* 33:290-299.
- Invernizzi P, Pasini S, Selmi C, Gershwin ME, and Podda M. 2009. Female Predominance and X Chromosome Defects in Autoimmune Diseases. *Journal of Autoimmunity* 33:12-16.
- Itin P, and Burger B. 2009. Mosaic Manifestations of Monogenic Skin Diseases. *Journal of the German Society of Dermatology* 7:744-748.
- Jablonka E, and Lamb M. 2008. Soft Inheritance: Challenging the Modern Synthesis. *Genetics and Molecular Biology* 31(2):389-395.
- Jaenisch R. 1997. DNA Methylation and Imprinting: why bother? *Trends in Genetics* 13(8):323-329.
- Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuissou J, Heredia F, Audigier P et al. . 2009. Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits. *PLoS Genetics* 5(6).
- Kabi A, Nickerson KP, Homer CR, and McDonald C. 2012. Digesting the Genetics of Inflammatory Bowel Disease: insights from studies of autophagy risk genes. *Inflammatory Bowel Disease* 18:782-792.
- Kaminsky Z, Petronis A, Wang S-C, Levine B, Ghaffar O, Floden D, and Feinstein A. 2008. Epigenetics of Personality Traits: An Illustrative Study of Identical Twins Discordant for Risk-Taking Behavior. *Twin Research and Human Genetics* 11(1):1-11.
- Kantlehner M, Kirchner R, Hartmann P, Ellwart JW, Alunni-Fabbroni M, and Schumacher A. 2011. A High-throughput DNA Methylation Analysis of a Single Cell. *Nucleic Acids Research* 39(7):e44-e52.
- Karaiskos C, Hudspith BN, Elliott T, Rayment NB, Avgousti V, Brostoff J, and Sanderson JD. 2011. Defective Macrophage Function in Crohn's Disease: Role of Alternatively Activated Macrophages in Inflammation. *American Gastroenterological Association Meeting*. Chicago, IL. p Sa1774.
- Karatzas PS, Mantzaris GJ, Safioleas M, and Gazouli M. 2014. DNA Methylation Profile of Genes Involved in Inflammation and Autoimmunity in Inflammatory Bowel Diseases. *Medicine* 93(28):309-315.
- Karlinger K, Gyorke T, Mako E, and Tarjan Z. 2000. The Epidemiology and the Pathogenesis of Inflammatory Bowel Disease. *European Journal of Radiology* 35:154-167.
- Kaser A, and Blumberg RS. 2011. Autophagy, Microbial Sensing, Endoplasmic Reticulum Stress, and Epithelial Function in Inflammatory Bowel Disease. *Gastroenterology* 140:1738-1747.
- Kiani J, and Rassoulzadegan M. 2013. A Load of Small RNAs in the Sperm - how many bits of hereditary information? *Cell Research* 23:18-19.
- Kidd K, Pakstis A, Speed W, and Kidd J. 2004. Understanding Human DNA Sequence Variation. *Journal of Heredity* 95:406-420.
- Kierszenbaum AL. 2002. Histology and Cell Biology: an introduction to pathology. St. Louis: Mosby.
- Kim SW, Kim ES, Moon CM, Kim TI, Kim WH, and Cheon JH. 2012. Abnormal Genetic and Epigenetic Changes in Signal Transducer and Activator of Transcription 4 in the Pathogenesis of Inflammatory Bowel Diseases. *Digestive Diseases and Science* 57:2600-2607.

- Kim Y-G, Shaw MH, Warner N, Park J-H, Chen F, Ogura Y, and Nunez G. 2011. Cutting Edge: Crohn's Disease-Associated NOD2 Mutation Limits Production of Proinflammatory Cytokines to Protect the Host from *Enterococcus faecalis*-Induced Lethality. *Journal of Immunology* 187:2849-2852.
- Klein A, and Eliakim R. 2010. Non Steroidal Anti-Inflammatory Drugs and Inflammatory Bowel Disease. *Pharmaceuticals* 3(4):1084-1092.
- Klein W, Tromm A, Folwaczny C, Hagedorn M, Duerig N, Epplen J, Schmiegel W, and Griga T. 2005. A Polymorphism of the Bactericidal/Permeability Increasing Protein (BPI) Gene Is Associated With Crohn's Disease. *Journal of Clinical Gastroenterology* 39:282-283.
- Kotlyar DS, Shum M, Hsieh J, Blonski W, and Greenwald DA. 2014. Non-pulmonary Allergic Diseases and Inflammatory Bowel Disease: a qualitative review. *World Journal of Gastroenterology* 20(32):11023-11032.
- Koziel JE, Fox MJ, Steding CE, Sprouse AA, and Herbert B-S. 2011. Medical Genetics and Epigenetics of Telomerase. *Journal of Cellular and Molecular Medicine* 15(3):457-467.
- Kristensen LS, Mikeska T, Krypuy M, and Dobrovic A. 2008. Sensitive Melting Analysis after Real Time- Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. *Nucleic Acids Research* 36(7):342.
- Kuratomi G, Iwamoto K, Bundo M, Kusumi I, Kato N, Iwata N, Ozaki N, and Kato T. 2008. Aberrant DNA Methylation Associated with Bipolar Disorder Identified from Discordant Monozygotic Twins. *Molecular Psychiatry* 13:429-441.
- Langmuir A. 1985. Medicine: Is Thucydides Syndrome Back? *New England Journal of Medicine* 313:1027-1030.
- Lawes D, Pearson T, SenGupta S, and Boulos P. 2005. The Role of *MLH1*, *MSH2* and *MSH6* in the Development of Multiple Colorectal Cancers. *British Journal of Cancer* 93:472-477.
- Le Souef PN, Candelaria P, and Goldblatt J. 2006. Evolution and Respiratory Genetics. *European Respiratory Genetics* 28:1258-1263.
- Ledford H. 2008. Disputed Definitions. *Nature* 455(7216):1023-1028.
- Lee G, and Buchman AL. 2009. DNA-driven Nutritional Therapy of Inflammatory Bowel Disease. *Nutrition* 25:885-891.
- Leeb S, Vogl D, Gunckel M, Kiessling S, Falk W, Goke M, Scholmerich J, Gelbmann CM, and Rogler G. 2003. Reduced Migration of Fibroblasts in Inflammatory Bowel Disease: Role of Inflammatory Mediators and Focal Adhesion Kinase. *Gastroenterology* 125:1341-1354.
- Lewis GJ, Massey DCO, Zhang H, Bredin F, Tremelling M, Lee JC, Berzuini C, and Parkes M. 2011. Genetic Association Between NLRP3 Variants and Crohn's Disease Does Not Replicate in a Large UK Panel. *Inflammatory Bowel Disease* 17:1387-1391.
- Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A et al. . 2007. Novel Crohn Disease Locus Identified by Genome-Wide Association Maps to a Gene Desert on 5p13.1 and Modulates Expression of PTGER4. *PLoS Genetics* 3(4):e58.
- Lin L-J, Zhang Y, Lin Y, Jin Y, and Zheng C-Q. 2014. Identifying Candidate Genes for Discrimination of Ulcerative Colitis and Crohn's Disease. *Molecular Biology Reports* 1033(014).
- Lin Z, Hegarty J, Cappel J, Yu W, Chen X, Faber P, Wang Y, Kelly A, Poritz L, Peterson B et al. . 2010. Identification of Disease-associated DNA Methylation in Intestinal Tissues from Patients with Inflammatory Bowel Disease. *Clinical Genetics* 80:59-67.

- Lin Z, Hegarty JP, Yu W, Cappel JA, Chen X, Faber PW, Wang Y, Poritz LS, Fan J-B, and Koltun WA. 2012. Identification of Disease-Associated DNA Methylation in B Cells from Crohn's Disease and Ulcerative Colitis Patients. *Digestive Disease Science* 57:3145-3153.
- Low JH, F.A. Williams, X. Yang, S. Cullen, J. Colley, K.L. Ling, A. Armuzzi, T. Ahmad, M.J. Neville, B.M. Dechairo, R. Walton, N.J. Lench, and D.P. Jewell. 2004. Inflammatory Bowel Disease is Linked to 19p13 and associated with ICAM-1. *Inflammatory Bowel Disease* 10:173-181.
- Magin WS, Van Kruiningen H, J., and Colombel J-F. 2013. Immunohistochemical Search for Viral and Bacterial Antigens in Crohn's Disease. *Journal of Crohn's and Colitis* 7:161-166.
- Mahmoudi M. 2014. Immunology Made Ridiculously Simple. Miami: MedMaster, Inc.
- Marcil V, Levy E, Mack DR, Kumar V, Israel DM, Deslandres C, Faure C, Sinnett D, Seidman EG, and Amre DK. 2011. The PTPN2 Gene and Pediatric Crohn's Disease. Association with Multiple Markers and Enhanced Expression in Intestinal Inflammation. *American Gastroenterological Association Meeting*. Chicago, IL. p Sa1271.
- Markowitz J, Leleiko NS, Keljo DJ, Mack DR, Evans J, Kappelman M, Oliva-Hemker M, Otley AR, Pfeifferkorn MD, Rosh JR et al. . 2011. The Prometheus Crohn's Prognostic Test Does Not Reliably Predict Complicated Crohn's Disease in Children. *American Gastroenterological Association Meeting*. Chicago, IL. p 933.
- Marquez L, Shen C, Machiels K, Perrier C, Ballet V, Organe S, Ferrante M, Henckaerts L, Rutgeerts PJ, Vermeire S et al. . 2011. Role of Haptoglobin in Susceptibility of IBD and in Triggering Murine Colitis. *American Gastroenterological Association Meeting*. Chicago, IL. p 114.
- Martos SN, Tang W-y, and Wang Z. 2015. Elusive Inheritance: transgenerational effects and epigenetic inheritance in human environmental disease. *Progress in Biophysics and Molecular Biology* 118:44-54.
- Mathieu N, Dumestre-Perard C, Kheav VD, Bichard P, Cesbron J-Y, and Bonaz BL. 2011. ASCA/OMP Are Serological Markers of Crohn's Disease Disability and Prognosis. *American Gastroenterological Association Meeting*. Chicago, IL. p Mo1979.
- McDermott E, Ryan EJ, Tosetto M, Gibson D, Burrage J, Keegan D, Byrne K, Crowe E, Sexton G, Manole K et al. . 2016. DNA Methylation Profiling in Inflammatory Bowel Disease Provides New Insights into Disease Pathogenesis. *Journal of Crohn's and Colitis*(1-10).
- McDonald JH. 2014. Handbook of Biological Statistics: 3rd Edition. Baltimore, MD: Sparky House Publishing. 305 p.
- McGovern DPB, Jones MR, Taylor KD, Marcianti K, Yah X, Dubinsky M, Ippoliti A, Vasilias E, Berel D, Derkowski C et al. . 2010. *Fucosyltransferase 2 (FUT2)* Non-secretor Status is Associated with Crohn's Disease. *Human Molecular Genetics* 19(17):3468-3476.
- Menoza JL, Medrano LM, Dema B, Taxonera C, Diaz-Rubio M, Urcelay E, and Nunez C. 2011. TLR9 and IL23R Genetic Interaction Modulates Crohn's Disease Susceptibility. *American Gastroenterological Association Meeting*. Chicago, IL. p Su1737.
- Metchnikoff E. 1905. Immunity in Infective Diseases: Cambridge University Press.
- Murdoch TB, Bernstein CN, El-Gabalawy H, Pathan S, Stempak JM, Sargent M, Xu W, Elias B, and Silverberg MS. 2011. Analysis of Inflammatory Bowel Disease-Associated Genetic



- Variants in a Healthy Manitoban First Nations Cohort. *American Gastroenterological Association Meeting*. Chicago, IL. p Sa1274.
- Myles S, Davison D, Barrett J, Stoneking M, and Timpson N. 2008. Worldwide Population Differentiation at Disease-Associated SNPs. *BMC Medical Genomics* 1:22.
- Nakagome S, Mano S, Kozlowski L, Bujinicki JM, Shibata H, Fukumaki Y, Kidd JR, Kidd KK, Kawamura S, and Oota H. 2012. Crohn's Disease Risk Alleles on the *NOD2* Locus Have Been Maintained by Natural Selection on Standing Variation. *Molecular Biology Evolution* 29(6):1569-1585.
- Nakagome S, Takeyama Y, Mano S, Matsui T, Kawamura S, and Oota H. 2010. Population-Specific Susceptibility to Crohn's Disease and Ulcerative Colitis; dominant and recessive relative risks in the Japanese population. *Annals of Human Genetics* 74:126-136.
- Nakazawa DJ. 2008. The Autoimmune Epigenic: bodies gone haywire in a world out of balance - and the cutting-edge science that promises hope. New York, London, Toronto, Sydney: Simon & Schuster.
- Newman WG, Zhang Q, Liu X, Amos CI, and Siminovich KA. 2009. Genetic Variants in IL-23R and ATG16L1 Independently Predispose to Increased Susceptibility to Crohn's Disease in a Canadian Population. *Journal of Clinical Gastroenterology* 43:444-447.
- Ng SC, Tsoi KKF, Kamm MA, Xia B, Wu J, Chan FKL, and Sung JJY. 2012. Genetics of Inflammatory Bowel Disease in Asia: Systematic Review and Meta-analysis. *Inflammatory Bowel Disease* 18:1164-1176.
- Nguyen M, Bradford K, Huang BL, Haritunians T, Fleshner P, Himmelfarb E, Landers CJ, Vasiliauskas EA, Ippoliti A, Melmed GY et al. . 2011. Genetic Predictors of Cytomegalovirus Infection in Patients With Inflammatory Bowel Disease. *American Gastroenterological Association Meeting*. Chicago, IL. p SA1284.
- Nielsen CH, Larsen A, and Nielsen AL. 2016. DNA Methylation Alterations in Response to Prenatal Exposure of Maternal Cigarette Smoking: a persistent epigenetic impact on health from maternal lifestyle? *Archives of Toxicology* 90:231-245.
- NIH. 2008. NIH Scientists Discover Crucial Control in Long-Lasting Immunity.
- Nimmo ER, Prendergast JG, Aldhouse MC, Kennedy NA, Henderson P, Drummond HE, Ramsahoye BH, Wilson DC, Semple CA, and Satsangi J. 2012. Genome-wide Methylation Profiling in Crohn's Disease Identifies Altered Epigenetic Regulation of Key Host Defence Mechanisms Including the Th17 Pathway. *Inflammatory Bowel Disease* 18(5):889-899.
- Nimmo ER, Stevens C, Phillips AE, Smith A, Drummond HE, Noble CL, Quail M, Davies G, Aldhouse MC, Wilson DC et al. . 2011. TLE1 Modifies the Effects of NOD2 in the Pathogenesis of Crohn's Disease. *Gastroenterology* 141:972-983.
- O'Hagan HM. 2014. Chromatin Modifications During Repair of Environmental Exposure-Induced DNA Damage: a potential mechanism for stable epigenetic alterations. *Environmental and Molecular Mutagenesis* 55:278-291.
- Ohmen JD, H. Yang, K.K. Yamamoto, H. Zhao, Y. Ma, L.G. Bentley, Z. Huang, S. Gerwehr, S. Pressman, C. McElree, S. Targan, J.I. Rotter, and N. Fischel-Ghodsian. 1996. Susceptibility Locus for Inflammatory Bowel Disease on Chromosome 16 has a Role in Crohn's Disease, but not in Ulcerative Colitis. *Human Molecular Genetics* 5:1679-1683.
- Olson P, Hames C, Benenson A, and Genovese E. 1996. The Thucydides syndrome: ebola deja vu? (or ebola reemergent). *Emerging Infectious Diseases* 2:155-156.

- Oota H, Pakstis AJ, Bonne-Tamir B, Goldman D, Grigorenko E, Kajuna SLB, Karom NJ, Kungulilo S, Lu R-B, Odunsi K et al. . 2004. The Evolution and Population Genetics of the *ALDH2* Locus: random genetic drift, selection, and low levels of recombination. *Annals of Human Genetics* 68:93-109.
- Padilla JC, Chaparro PE, Molina K, Arevalo-Herrera M, and Herrera S. 2015. Is There Malaria Transmission in Urban Settings in Columbia? *Malaria Journal* 14:453.
- Palmieri O, Latiano A, Maglietta R, Palumbo O, Corritore G, Latiano T, Martino G, Carella M, Ancona N, Bossa F et al. . 2011. Dissection of the Crohn's Disease Transcriptome of 71 Loci Using Genome-Wide Microarrays. *American Gastroenterological Association Meeting*. Chicago, IL. p SA1285.
- Panchenko PE, Voisin S, Jouin M, Jouneau L, Prezelin A, Lecoutre S, Breton C, Jammes H, Junien C, and Gabory A. 2016. Expression of Epigenetic Machinery Genes is Sensitive to Maternal Obesity and Weight Loss in Relation to Fetal Growth in Mice. *Clinical Epigenetics* 8(22):1-19.
- Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D et al. . 2007. Sequence Variants in the Autophagy Gene IRGM and Multiple Other Replicating Loci Contribute to Crohn's Disease Susceptibility. *Nature Genetics* 39(7):830-832.
- Peeters M, Nevens H, Baert F, and al. e. 1996. Familial Aggregation in Crohn's Disease: increased age-adjusted risk and concordance in clinical characteristics. *Gastroenterology* 111:597-603.
- Pena AS. 2006. Contribution of Genetics to a New Vision in the Understanding of Inflammatory Bowel Disease. *World Journal of Gastroenterology* 12:4784-4787.
- Pepoyan A, Balayan, Marine, Manvelyan A, Marutan S, Minasbekyan L, Hovnanyan K, and Tsaturyan V. 2015. DNA Fluorescence Parameters and Methylation Levels of Gut Commensal *Escherichia coli* from Crohn's Disease Patients. *Biophysical Journal* 106(2):726.
- Peter I, Mitchell AA, Ozelius L, Erazo M, Hu J, Doheny D, Abreu MT, Present DH, Ullman T, Benkov K et al. . 2011. Evaluation of 22 genetic variants with Crohn's Disease risk in the Ashkenazi Jewish population: a case-control study. *BMC Medical Genetics* 12:1-9.
- Petronis A, and Petroniene R. 2000. Epigenetics of Inflammatory Bowel Diseases. *Gut* 47:302-306.
- Pewsner D, Battaglia M, Minder C, Marx A, Bucher H, and Egger M. 2004. Ruling a Diagnosis in or out with "SpPin" and "SnNOut": a word of caution. *British Medical Journal (clinical research ed)* 329(7459):209-213s.
- Pierce ES. 2009. Possible Transmission of *Mycobacterium avium* subspecies paratuberculosis Through Potable Water: lessons from an urban cluster of Crohn's disease. *Gut Pathogens* 1:1-17.
- Plotkin S. 2005. Vaccines: past, present and future. *Nature Medicine* 11(4 Suppl):S5-11.
- Pugliatti M, Sotgiu S, and Solinas G. 2001. Multiple Sclerosis Epidemiology in Sardinia: evidence for a true increasing risk. *Acta Neurology Scandinavia* 103:20-26.
- Rafehi H, El-Osta A, and Karagiannis TC. 2011. Genetic and Epigenetic Events in Diabetic Wound Healing. *International Wound Journal* 8:12-21.
- Ragunathan K, Jih G, and Moazed D. 2014. Epigenetic Inheritance Uncoupled from Sequence-Specific Recruitment. *Scienceexpress*(20 November 2014):1-13.
- Rando OJ. 2012. Daddy Issues: Paternal Effects on Phenotype. *Cell* 151:702-708.

- Rapozo DC, Bernadazzi C, and de Souza HSP. 2017. Diet and Microbiota in Inflammatory Bowel Disease: the gut in disharmony. *World Journal of Gastroenterology* 23(12):2124-2140.
- Reece K. 2012. DNA Methylation Mechanisms and Analysis Methods to Study this Key Epigenetic Control. *Promega Webinar*.
- Reece K, Mook J, and Storts D. 2012. A Real-Time Dye-Based PCR Assay for Assessing Bisulfite Conversion Efficiency.
- Retief F, and Cilliers L. 1998. The Epidemic of Athens, 430-426 BC. *South African Medical Journal* 88:50-53.
- Rieder F, and Fiocchi C. 2008. Intestinal Fibrosis in Inflammatory Bowel Disease - Current Knowledge and Future Perspectives. *Journal of Crohn's and Colitis* 2(4):279-290.
- Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW et al. . 2007. Genome-wide Association Study Identifies New Susceptibility Loci for Crohn Disease and Implicates Autophagy in Disease Pathogenesis. *Nature Genetics* 39(5):596-604.
- Risques RA, Rabinovitch PS, and Brentnall TA. 2006. Cancer Surveillance in Inflammatory Bowel Disease: new molecular approaches. *Current Opinion in Gastroenterology* 22:382-390.
- Rodriquez-Bores L, G.C. Fonseca, M.A. Villeda, and J.K. Yamamoto-Furusho. 2007. Novel Genetic Markers in Inflammatory Bowel Disease. *World Journal of Gastroenterology* 13(42):5560-5570.
- Rose M, Hildebrandt M, Fliege H, Seibold S, Mönnikes H, and Klapp BF. 2002. T-Cell Immune Parameters and Depression in Patients With Crohn's Disease. *Journal of Clinical Gastroenterology* 34(1):40-48.
- Rosenstiel PC, Barann M, Franke A, Stade B, Thomsen I, Schilhabel MB, Sheth V, Lee C, Klostermeier UC, McKernan K et al. . 2011. Whole Genome SEquence of a Crohn Disease Trio - A Paradigm for Individualized Disease Etiology Discovery. *American Gastroenterological Association Meeting*. Chicago, IL. p 113.
- Russell RK, H.E. Drummond, E.R. Nimmo, N.H. Anderson, C.L. Noble, D.C. Wilson, P.M. Gillett, P. McGrogan, K. Hassan, L.T. Weaver, W.M. Bisset, G. Mahdi, and J. Satsangi. 2006. Analysis of the Influence of OCTN1/2 Variants within the IBD5 Locus on Disease Susceptibility and Growth Indices in Early Onset Inflammatory Bowel Disease. *Gut* 55:1114-1123.
- Ryan JD, Silverberg MS, Xu W, Walker JR, Graff LA, Ediger J, Miller N, Rogala L, Sargent M, Carr R et al. . 2011. New Insights Into the Genetic Basis for Variation in Crohn's Disease (CD) Phenotype in a Population-Based Study. *American Gastroenterological Association Meeting*. Chicago, IL. p Tu1269.
- Sadler T, Bhasin JM, Xu Y, Bamholz-Sloan J, Chen Y, Ting AH, and Stylianou E. 2017. Genome-Wide Analysis of DNA Methylation and Gene Expression Defines Molecular Characteristics of Crohn's Disease-Associated Fibrosis. *Clinical Epigenetics* 8(30).
- Sahasranaman S, Howard D, and Roy S. 2008. Clinical Pharmacology and Pharmacogenetics of Thiopurines. *European Journal of Clinical Pharmacology* 64(8):753-767.
- Sasikala M, Reddy DN, Pratap N, Sharma SK, Balkumar PR, Sekeran A, Banerjee R, and Reddy DB. 2009. Absence of *Mycobacterium avium* ss paratuberculosis-specific IS900 Sequence in Intestinal Biopsy Tissues of Indian Patients with Crohn's Disease. *Indian Journal of Gastroenterology* 28(5):169-174.

- Sato F, Tsuchiya S, Meltzer SJ, and Shimizu K. 2011. MicroRNAs and Epigenetics. *FEBS Journal* 278:1598-1609.
- Satsangi J, M. Parkes, D.P. Jewell, and J.I. Bell. 1998. Genetics of Inflammatory Bowel Disease. *Clinical Science (London)* 94:473-478.
- Satsangi J, M. Parkes, E. Louis, L. Hashimoto, N. Kato, K. Welsh, J.D. Terwilliger, G.M. Lathrop, J.I. Bell, and D.P. Jewell. 1996. Two Stage Genome-Wide Search in Inflammatory Bowel Disease Provides Evidence for Susceptibility Loci on Chromosomes 3, 7 and 12. *Nature Genetics* 14:199-202.
- Scharl M, Paul G, Weber A, Jung BC, Docherty MJ, Hausmann M, Rogler G, Barrett KE, and McCole DF. 2009. Protection of Epithelial Barrier Function by the Crohn's Disease Associated Gene Protein Tyrosine Phosphatase N2. *Gastroenterology* 137:2030-2040.
- Scharl M, Wojtal KA, Becker HM, Fischbeck A, Arikkat JM, Pesch T, Kellermeier S, Weber A, Loessner MJ, Frei P et al. . 2011a. The Crohn's Disease Associated Variant of the Protein Tyrosine Phosphatase Non-Receptor Type 2 Gene Affects Cellular Responses to Invading *Listeria Monocytogenes*. *American Gastroenterological Association Meeting*. Chicago, IL. p Su1789.
- Scharl M, Wojtal KA, Fischbeck A, Arikkat JM, Pesch T, Kellermeier S, Vavricka SR, Fried M, McCole DF, and Rogler G. 2011b. The Crohn's Disease Candidate Gene, Protein Tyrosine Phosphatase Non-Receptor Type 2, Regulates Muramyl dipeptide-Induced NOD2-Dependent Effects in Human Monocytes and Fibroblasts. *American Gastroenterological Association Meeting*. Chicago, IL. p Su1744.
- Schoenborn J, and Wilson C. 2007. Regulation of Interferon-Gamma During Innate and Adaptive Immune Responses. *Advances in Immunology* 96:41-101.
- Serbati N, Senhaji N, Diakite B, Badre W, and Nadifi S. 2014. IL23R and ATG16L1 Variants in Moroccan Patients with Inflammatory Bowel Disease. *BMC Research Notes* 7(570).
- Shapiro R, Braverman B, Louis JB, and Servis RE. 1973. Nucleic Acid Reactivity and Conformation: II. Reaction of Cytosine and Uracil with Sodium Bisulfite. *The Journal of Biological Chemistry* 248:4060-4064.
- Shin A-R, Kim H-J, Cho SN, Collins MT, Manning EJB, Naser SA, and Shin SJ. 2009. Identification of Seroreactive Proteins in the Culture Filtrate Antigen of *Mycobacterium avium* ssp. paratuberculosis Human Isolates to Sera from Crohn's Disease Patients. *FEMS Immunology & Medical Microbiology* 58:128-137.
- Sibartie S, Scully P, Keohane J, O'Neill S, O'Mahony J, O'Hanlon D, Kirwan WO, O'Mahony L, and Shanahan F. 2010. *Mycobacterium avium* subsp. paratuberculosis (MAP) as a Modifying Factor in Crohn's Disease. *Inflammatory Bowel Disease* 16:296-304.
- Siegel CA, Fleshner P, Siegel LS, Rotter JJ, Dubinsky M, Vasiliauskas EA, Ippoliti A, Shih DQ, Melmed GY, Landers CJ et al. . 2011. Predicting Crohn's Disease Post-Operative Recurrence Using Clinical, Endoscopic, Serologic and Genetic Factors. *American Gastroenterological Association Meeting*. Chicago, IL. p 935.
- Simmons JD, C. Mullighan, K.I. Welsh, and D.P. Jewell. 2000. Vitamin D Receptor Gene Polymorphism: Association with Crohn's Disease Susceptibility. *Gut* 47:211-214.
- Snoeks L, Weber CR, Wasland K, Turner JR, Vainder C, Qi W, and Savkovic SD. 2009. Tumor Suppressor FOXO3 Participates in the Regulation of Intestinal Inflammation. *Laboratory Investigation* 89:1053-1062.
- Sokol H, Vasquez N, Hoyeau-Idrissi N, Seksik P, Beaugerie L, Lavergne-Slove A, Pochart P, and Marteau P. 2010. Crypt Abscess-associated Microbiota in Inflammatory Bowel

- Disease and Acute Self-limited Colitis. *World Journal of Gastroenterology* 16(5):583-587.
- Sompayrac L. 2012. How the Immune System Works. Chichester, West Sussex: Wiley-Blackwell.
- Spitale R, Tsai M, and Chang H. 2011. RNA Templating the Epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* 6(5):539-543.
- Stanislawowski M, Wierzbicki PM, Golab A, Adrych K, Kartanowicz D, Wypych J, Godlewski J, Smoczynski M, and Kmiec Z. 2009. Decreased Toll-Like Receptor-5 (TLR-5) Expression in the Mucosa of Ulcerative Colitis Patients. *Journal of Physiology and Pharmacology* 60(4):71-75.
- Stenvinkel P, Karimi M, Johansson S, Axelsson J, Suliman M, and Lindholm B. 2007. Impact of Inflammation on Epigenetic DNA Methylation - a novel risk factor for cardiovascular disease? *Journal of Internal Medicine* 261(5):488-499.
- Sura R, Colombel J, and Van Kruiningen H. 2011. Lymphatics, Tertiary Lymphoid Organs and the Granulomas of Crohn's Disease: an immunohistochemical study. *Alimentary Pharmacology Therapy*(33):930-939.
- Suskind D, Brittnacher M, Wahbeh G, Shaffer M, Haden H, Qin X, Singh N, Damman C, Hager K, Nielson H et al. . 2015. Fecal Microbial Transplant Effect on Clinical Outcomes and Fecal Microbiome in Active Crohn's Disease. *Inflammatory Bowel Disease* 21(3):556-563.
- Sventoraityte J, Zvirbliene A, Franke A, Kwiatkowski R, Kiudelis G, Kupcinskas L, and Schreiber S. 2010. NOD2, IL23R and ATG16L1 Polymorphisms in Lithuanian Patients with Inflammatory Bowel Disease. *World Journal of Gastroenterology* 16(3):359-364.
- Swarbrick ET, Farrokyar F, and Irvine EJ. 2001. A Critical Review of Epidemiological Studies in Inflammatory Bowel Disease. *Scandinavian Journal of Gastroenterology* 36:2-15.
- Sykora J, I. Subrt, P. Dedek, S. Konrad, J. Schwarz, V. Machalova, J. Varvarovska, P. Pazdiora, O. Pozler, and F. Stozicky. 2006. Cytokine Tumor Necrosis Factor-alpha A Promotor Gene Polymorphism at Position – 308 G->A and Pediatric Inflammatory Bowel Disease: Implications in Ulcerative Colitis and Crohn's Disease. *Journal of Pediatric Gastroenterology and Nutrition* 42(5):479-487.
- Thiede DA. 1998. Maternal Inheritance and Its Effect on Adaptive Evolution: a quantitative genetic analysis of maternal effects in a natural plant population. *Evolution* 52(4):998-1015.
- Thompson NP, Driscoll R, Pounder RE, and Wakefield AJ. 1996. Genetics Versus Environment in Inflammatory Bowel Disease: results of a British twin study. *British Medical Journal* 312:95-96.
- Tollefsbol T. 2010. Handbook of Epigenetics: The New Molecular and Medical Genetics: Academic Press.
- Török HP, J. Glas, L. Tonenchi, P. Lohse, B. Muller-Myhshok, O. Limbersky, C. Neugebauer, F. Schnitzler, J. Seiderer, C. Tillack, S. Brand, G. Brunneler, P. Jagiello, J.T. Epplen, T. Griga, W. Klein, U. Schiemann, M. Folwaczny, T. Ochsenkuhn, Folwaczny, C. 2005. Polymorphisms in the DLG5 and OCTN Cation Transporter Genes in Crohn's Disease. *Gut* 54:1421-1427.
- Tuomilehto j, Karvonen M, and Pitkaniemi J. 1999. Record-high Incidence of Type I (insulin-dependent) Diabetes Mellitus in Finnish Children. *Diabetologia* 42:655-660.

- Tysk C, Lindberg E, Jarnerot G, and Floderus-Myrhed B. 1988. Ulcerative Colitis and Crohn's Disease in an Unselected Population of Monozygotic and Dizygotic Twins: a study of heritability and the influence of smoking. *Gut* 29:990-996.
- Varner AE. 2002. The Increase in Allergic Respiratory Diseases: Survival of the Fittest? *Chest* 121:1308-1316.
- Varona L, Munilla S, Casellas J, Moreno C, and Altarriba J. 2015. Consequences of Paternally Inherited Effects on the Genetic Evaluation of Maternal Effects. *Genetics Selection Evolution* 47(63):1-13.
- Velasquez-Manoff M. 2012. An Epidemic of Absence: A New Way of Understanding Allergies and Autoimmune Diseases. New York: Scribner.
- Ventham NT, Kennedy NA, Nimmo ER, and Satsangi J. 2013. Beyond Gene Discovery in Inflammatory Bowel Disease: The Emerging Role of Epigenetics. *Gastroenterology* 145:293-308.
- Vermiere S, M. Pierik, T. Hlavaty, G. Claessens, N. van Schuerbeeck, S. Joossens, M. Ferrante, L. Henckaerts, M. Bueno de Mesquita, R. Vlietinck, and P. Rutgeerts. 2005. Association of Organic Cation Transporter Risk Haplotype with Perianal Penetrating Crohn's Disease but not with Susceptibility to IBD. *Gastroenterology* 125:1845-1853.
- Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, and Martienssen RA. 2002. Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science* 297:1833-1837.
- Waddington C. 1942. The Epigenotype. *Endeavor* 1:18-20.
- Wade MJ. 2001. Maternal Effect Genes and the Evolution of Sociality in Haplo-Diploid Organisms. *Evolution* 55(3):453-458.
- Wagner J, Glas J, Seiderer J, Goke B, Ochsenkuhn, Diegelmann J, and Brand S. 2011. Variants in the *PTPN2* Gene are Associated with Susceptibility to Both Crohn's Disease and Ulcerative Colitis in the German Population. *American Gastroenterological Association Meeting*. Chicago, IL. p S1291.
- Walsh C. 2006. Chapter 5 - Protein Methylation. *Posttranslational modification of proteins: expanding nature's inventory*. Roberts and Co. Publishers.
- Walzer N, and Buchman AL. 2010. Development of Crohn's Diseases in Patients With Intestinal Failure: A Role for Bacteria? *Journal of Clinical Gastroenterology* 00:000-000.
- Waterman M, Xu W, Stempak JM, Milgrom R, Bernstein CN, Griffiths AM, Greenberg GR, Steinhart AH, and Silverbert MS. 2011. Distinct and Overlapping Genetic Loci in Crohn's Disease and Ulcerative Colitis: Correlations with Pathogenesis. *Inflammatory Bowel Disease* 17(9):1936-1943.
- Wojdacz TK, Hansen LL, and Dobrovic A. 2008. A New Approach to Primer Design for the Control of PCR Bias in Methylation Studies. *BMC Research Notes* 1(54).
- Woof J, and Burton D. 2004. Human Antibody-Fc Receptor Interactions Illuminated by Crystal Structures. *Nature Reviews Immunology* 4(2):89-99.
- [http://www.gen.tcd.ie/molpopgen/link\\_files/Ireland\\_mtDNA.xls](http://www.gen.tcd.ie/molpopgen/link_files/Ireland_mtDNA.xls).
- <http://www.wikipedia.org/>.
- Zaahl M, Winter T, Warnich L, and Kotze M. 2005. Analysis of the Three Common Mutations in the *CARD15* gene (R702W, G908R and 1007fs) in South American Colored Patients with Inflammatory Bowel Disease. *Molecular Cell Probes* 19:278-281.

- Zelinkova Z, Stokkers PC, van der Linde K, Kuipers EJ, Peppelenbosch MP, and van der Woude CPJ. 2012. Maternal Imprinting and Female Predominance in Familial Crohn's disease. *Journal of Crohn's and Colitis* 6:771-776.
- Zheng CQ, G.Z. Hu, Z.S. Zeng, L.J. Line, and G.G. Gu. 2003. Progress in Searching for Susceptibility Genes for Inflammatory Bowel Disease by Positional Cloning. *World Journal of Gastroenterology* 9:1646-1656.
- Zouali H, S. Lesage, F. Merlin, J.P. Cezard, J.F. Colombel, J. Belaiche, S. Almer, C. Tysk, C. O'Morain, M. Gassuli, S. Christensen, Y. Finkel, R. Modigliani, C. Gower-Rousseau, J. Macry, M. Chamaillard, G. Thomas, and J.P. Hugot. 2003. CARD4/NOD1 is Not Involved in Inflammatory Bowel Disease. *Gut* 52:71-74.

## **APPENDICES**

### **APPENDIX 1 – IRB CONSENT FORM**

#### **RESEARCH CONSENT FORM**

##### **Differential Methylation of Immunological Genes: Epigenetics of Crohn's Disease**

You are being asked to join a research study. You are being asked to take part in this study because you have Crohn's disease or you are a healthy individual without Crohn's disease. You do not have to participate in this research study. The main purpose of research is to create new knowledge for the benefit of future patients and society in general. Research studies may or may not benefit the people who participate.

Research is voluntary, and you may change your mind at any time. There will be no penalty to you if you decide not to participate, or if you start the study and decide to stop early. Either way, you can still get medical care and services at the University of Kansas Medical Center (KUMC).

This consent form explains what you have to do if you are in the study. It also describes the possible risks and benefits. Please read the form carefully and ask as many questions as you need to, before deciding about this research.

You can ask questions now or anytime during the study. The researchers will tell you if they receive any new information that might cause you to change your mind about participating.



This research study will take place at the University of Kansas Medical Center (KUMC) with Dr. John Bonino as the researcher. Ms. Delisa Phillips, a graduate student at the University of Kansas, also is working on the project. About 200 people will be in the study at KUMC.

## **BACKGROUND**

The cells in your body contain deoxyribonucleic acid, or DNA for short. DNA is passed down from your parents. It carries the genes that determine how you look and how your body works. Differences in genes may help explain why a particular treatment is effective and safe in some people, but not in others. Differences in our genes may explain why some people get certain diseases, but others do not. The study of DNA is called genetic research.

## **PURPOSE OF THE STUDY**

The purpose of this project is to study the genetics of individuals with Crohn's Disease and compare them to individuals without the disease.

## **PROCEDURES**

If you agree to participate, you will be asked to fill out a short survey about your health. Then you will have DNA taken from your cheeks using a gentle swab. You will be assigned a code number. The code numbers of the samples will be used during the study in order to protect the confidentiality of your information. DNA from the cheek swabs will be extracted and analyzed. It will be stored at the Laboratory of Biological Anthropology at the University of Kansas. Your

DNA sample will only be used by the researchers on this study. The results of our study will be given to you at the conclusion of the study if you provide contact information.

## **RISKS**

There are no physical risks with this study. There is a small risk of your genetic information being shared in a way that is not allowed. However, the researchers will use a code to protect your identity, so this risk is very small.

## **GINA**

A federal law, called the Genetic Information Nondiscrimination Act (GINA), makes it illegal for health insurance companies, group health plans, and most employers to discriminate against you based on your genetic information. This law will protect you in the following ways:

- Health insurance companies and group health plans may not request your genetic information that we get from this research.
- Health insurance companies and group health plans may not use your genetic information when making decisions regarding your eligibility or premiums.
- Employers with 15 or more employees may not use your genetic information that we get from this research when making a decision to hire, promote, or fire you or when setting the terms of your employment. The GINA protections do not help you if you work for a company with less than 15 employees.

Be aware that this federal law does not protect you against genetic discrimination by companies that sell life insurance, disability insurance, or long-term care insurance.

## **BENEFITS**

You will not directly benefit from participating in this study. The information gained in the study will not be used for your personal health care. However, if you provide your address, we will send you a summary of what we learned at the end of the study. We hope that the genetic research will help researchers understand more about Crohn's disease and possibly lead to new treatments in the future.

## **ALTERNATIVES**

Participation in this research is entirely optional. You can decide not to participate and still receive services from the University of Kansas Medical Center.

## **COST/ PAYMENT**

There is no cost to be in the study. Participants in this study will not be paid.

## **IN THE EVENT OF INJURY**

No harm is expected from this study. However, if you have any problem or concern, please contact Dr. Bonino at (913) 588-6003.

If you think you have been harmed as a result of participating in research at the University of Kansas Medical Center (KUMC), you should contact the Director, Human Research Protection Program, Mail Stop #1032, University of Kansas Medical Center, 3901 Rainbow Blvd., Kansas City, KS 66160. Under certain conditions, Kansas state law or the Kansas Tort Claims Act may allow for payment to persons who are injured in research at KUMC

## **CONFIDENTIALITY AND PRIVACY AUTHORIZATION**

Efforts will be made to keep your personal information confidential. Researchers cannot guarantee absolute confidentiality. If the results of this study are published or presented in public, information that identifies you will be removed.

The privacy of your health information is protected by a federal law known as the Health Insurance Portability and Accountability Act (HIPAA). By signing this consent form, you are giving permission (“authorization”) for KUMC to use and share your health information for the purposes of this research study. If you decide not to sign the form, the cheek swab will not be taken and no information will be collected.

To do this research, the research team needs to view and collect health information that identifies you. If you are a patient at University of Kansas Medical Center, they may look in your medical record to find out if you qualify for the study. The study information includes your survey answers and the DNA sample from your cheek swab. Your DNA sample will only be labeled with a code. Only the KUMC researchers will know which code belongs to your name. Codes and names will be kept in a locked file or on a computer with a password. Research records might be seen by officials at KUMC and U.S. agencies who oversee research, but those officials must keep your information confidential.

By signing this form, you are giving Dr. Bonino to share coded information about you with other persons on the research team. Ms. Phillips and her team will do the genetic research at the Laboratory of Biological Anthropology at the University of Kansas. The outside groups that see

your research records will only have your code. They will not have your name or other identifying information. In this way, your information will not be disclosed by others and your privacy will remain protected. The researchers will not share information about you with anyone not specified above unless required by law or unless you give written permission.

Your genetic sample will be kept for the duration of this study. After that, it will be destroyed.

Permission granted on this date to use and disclose your information remains in effect for a period of five years. By signing this form, you give permission for the use and disclosure of your information for purposes of this study for the next five years.

You or your doctor will not get individual results about your genetic testing. The results of your genetic testing will not be put in your medical record.

#### QUESTIONS

Before you sign this form, Dr. Bonino should answer all your questions. You can talk to the researchers if you have any more questions, suggestions, concerns or complaints after signing this form. If you have any questions about your rights as a research subject, or if you want to talk with someone who is not involved in the study, you may call the Human Subjects Committee at (913) 588-1240. You may also write the Human Subjects Committee at Mail Stop #1032, University of Kansas Medical Center, 3901 Rainbow Blvd., Kansas City, KS 66160.

#### SUBJECT RIGHTS AND WITHDRAWAL FROM THE STUDY

You can stop being in the study at any time. Your decision to stop will not prevent you from getting treatment or services at KUMC.

Your permission to use and share your health information for this genetic research will not expire unless you cancel it. If you want to cancel your permission to use your information, please write to Dr. Bonino. The mailing address is Dr. John Bonino University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160. You may also write Dr. Bonino if you want your genetic sample to be destroyed. If you ask, the researchers will destroy your sample. They may use and share information that was gathered before they received your request.

### **CONSENT**

Dr. Bonino or the research team has given you information about this research study. They have explained what will be done and how long it will take. They explained any inconvenience, discomfort or risks that may be experienced during this study.

By signing this form, you say that you freely and voluntarily consent to participate in this research study. You have read the information and had your questions answered.

***You will be given a signed copy of the consent form to keep for your records.***

---

Print Participant's Name

---

Signature of Participant

---

Time

---

Date

---

Name of Person Obtaining Consent

---

Signature of Person Obtaining Consent

---

Date

## APPENDIX 2 – DATA COLLECTION WORKSHEET



Data Collection Worksheet

HSCL 19611

ID # \_\_\_\_\_

Date of Interview/Examination (MM/DD/YYYY): \_\_\_\_\_

1. Are you male or female?      ☐ 1 Male      ☐ 2 Female
2. What is your date of birth? MM/DD/YY \_\_\_\_\_
3. What is your ethnicity? \_\_\_\_\_
4. Where did you grow up? \_\_\_\_\_
5. Have you lived in primarily urban or rural environments?    ☐ Urban    ☐ Rural
6. Are you/were you a smoker? \_\_\_\_\_ Date of last cigarette \_\_\_\_\_
7. Do you drink alcohol? \_\_\_\_\_ Number of drinks/week \_\_\_\_\_
8. Do you have Crohn's Disease? \_\_\_\_\_
9. If you have Crohn's Disease, when were you diagnosed? \_\_\_\_\_
10. When did the symptoms first appear? \_\_\_\_\_
11. Do any foods make the symptoms worse (or the disease flare up if you are in remission)?  
Please list:  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_
12. Do you have any allergies? Please list all drug and food allergies:  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_
13. Had you experienced any stress around the time of the first onset of symptoms? If yes, please describe:  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_
14. Do you eat meat? \_\_\_\_\_



15. Do you have any family members with Crohn's Disease? \_\_\_\_\_

If yes, please list their relationship to you \_\_\_\_\_

Have you had any of these clinician-diagnosed illnesses?		Year of Diagnosis
Mark here for <b>Yes</b> → Leave blank for <b>NO</b> ;	<b>Yes</b>	
	↓	
Myocardial infarction (heart attack)	<input type="checkbox"/>	_____
Angina pectoris	<input type="checkbox"/>	_____
Confirmed angiography? <input type="checkbox"/> No <input type="checkbox"/> Yes		
Heart Disease	<input type="checkbox"/>	_____
Stroke (CVA)	<input type="checkbox"/>	_____
Deep vein thrombosis/Pulmonary embolism	<input type="checkbox"/>	_____
Cancer	<input type="checkbox"/>	_____
Type of Cancer:	<input type="checkbox"/>	_____
Fibrocystic/other benign breast disease	<input type="checkbox"/>	_____
Confirmed by breast biopsy? <input type="checkbox"/> No <input type="checkbox"/> Yes		
Confirmed by aspiration? <input type="checkbox"/> No <input type="checkbox"/> Yes		
Colon or rectal polyp (benign)	<input type="checkbox"/>	_____
Ulcerative colitis/Crohn's Disease	<input type="checkbox"/>	_____
Gastric or duodenal ulcer	<input type="checkbox"/>	_____
Barrett's Esophagus	<input type="checkbox"/>	_____
Gallstones	<input type="checkbox"/>	_____
Did you have symptoms? <input type="checkbox"/> No <input type="checkbox"/> Yes		
How diagnosed? <input type="checkbox"/> X-ray or ultrasound <input type="checkbox"/> Other		
Celiac Disease	<input type="checkbox"/>	_____
Diabetes mellitus	<input type="checkbox"/>	_____
Elevated cholesterol	<input type="checkbox"/>	_____
High blood pressure	<input type="checkbox"/>	_____
Lupus	<input type="checkbox"/>	_____
Anemia	<input type="checkbox"/>	_____
Kidney stones	<input type="checkbox"/>	_____
Asthma, doctor diagnosed	<input type="checkbox"/>	_____
Emphysema/Chronic Bronchitis, doctor diagnosis	<input type="checkbox"/>	_____
Pneumonia, x-ray confirmed	<input type="checkbox"/>	_____
Graves' Disease/Hyperthyroidism	<input type="checkbox"/>	_____
Hypothyroidism	<input type="checkbox"/>	_____

Back problems Please specify type:	<input type="checkbox"/>	_____
Rheumatoid arthritis, doctor diagnosis	<input type="checkbox"/>	_____
Other arthritis	<input type="checkbox"/>	_____
Depression, clinician diagnosis	<input type="checkbox"/>	_____
Other major illness or surgery	<input type="checkbox"/>	_____
Please specify: Date:		
Eye problems (such as iritis) Please specify type:	<input type="checkbox"/>	_____

## APPENDIX 3 – COUNT OF EACH VARIABLE FROM QUESTIONNAIRE

Table 23 Count of each variable from the Questionnaire broken down by Crohn's disease, ulcerative colitis, and phenotypically normal samples.

Variable	Crohn's disease	Ulcerative colitis	Control	Total
Gender: Male	7	5	15	27
Gender: Female	13	5	26	44
Age: 18-25	2	0	2	4
Age: 26-40	7	4	8	19
Age: 41-55	9	4	9	22
Age: 56+	2	2	22	26
Ethnicity: European	17	8	36	61
Ethnicity: African	1	0	4	5
Ethnicity: Asian	1	1	0	2
Ethnicity: Native American	1	1	1	3
Region: Midwest	12	6	33	51
Region: South	1	1	4	6
Region: West	2	0	0	2
Region: North	2	1	0	3
Region: Northeast	1	1	0	2
Region: Southwest	1	0	2	3
Region: Southeast	0	0	2	2
Region: Asia	1	1	0	2
Environment: Urban	10	5	27	42
Environment: Rural	10	5	14	29
Smoker: Yes	6	4	10	20
Smoker: No	14	6	31	51
Alcohol: Yes	8	3	16	27
Alcohol: No	12	7	25	44
Allergies: Yes	14	3	14	31
Allergies: No	6	7	27	40
Familial CD: Yes	5	0	2	7
Familial CD: No	15	10	39	64
Heart attack: Yes	0	0	5	5
Heart attack: No	20	10	36	66
Cancer: Yes	1	1	9	11
Cancer: No	19	9	32	60
Celiac disease: Yes	1	1	2	4
Celiac disease: No	19	9	39	67
High cholesterol: Yes	1	1	15	17
High cholesterol: No	19	9	26	54
Asthma: Yes	2	1	8	11

<b>Asthma: No</b>	18	9	33	60
<b>Rheumatoid arthritis: Yes</b>	3	0	4	7
<b>Rheumatoid arthritis: No</b>	17	10	37	64

## APPENDIX 4 – AMPLIFICATION PLOTS AND MELT CURVES

### AMPLIFICATION PLOTS AND MELT CURVES OF *NOD2BC*

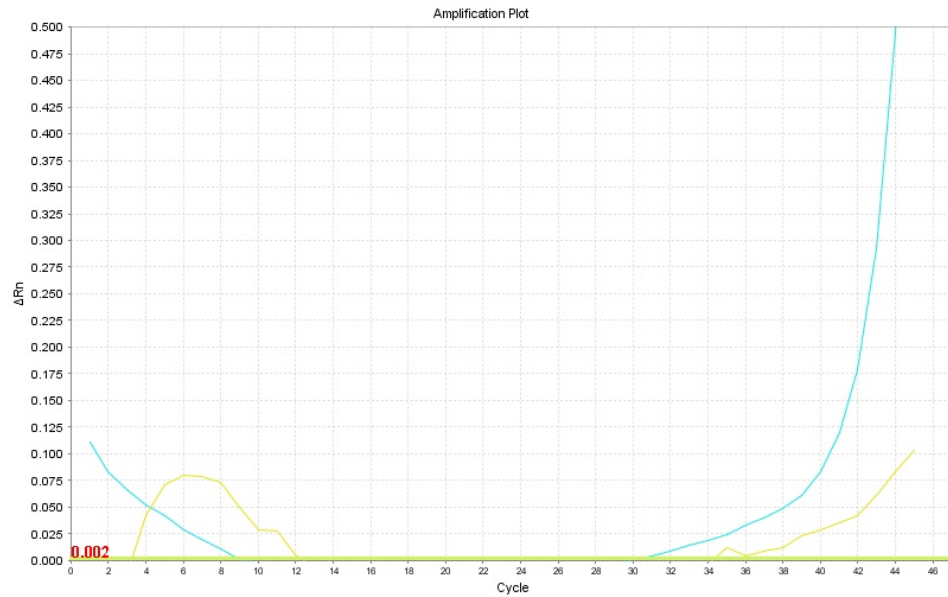


Figure 11 Amplification plot of *NOD2bc* for Crohn's disease patients.

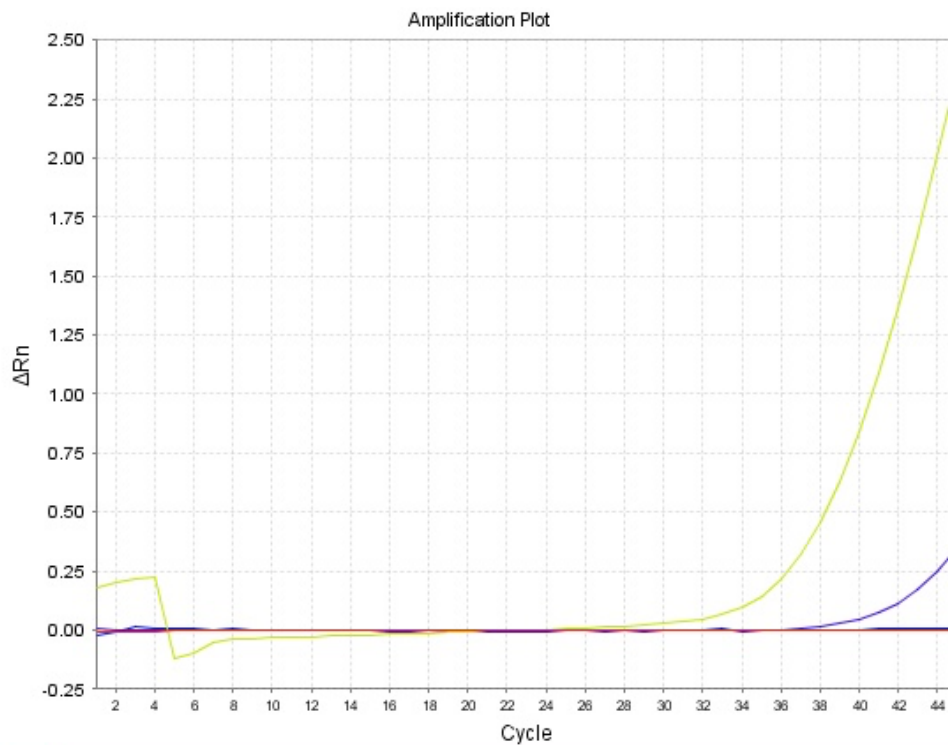


Figure 12 Amplification plot of *NOD2bc* for ulcerative colitis samples.

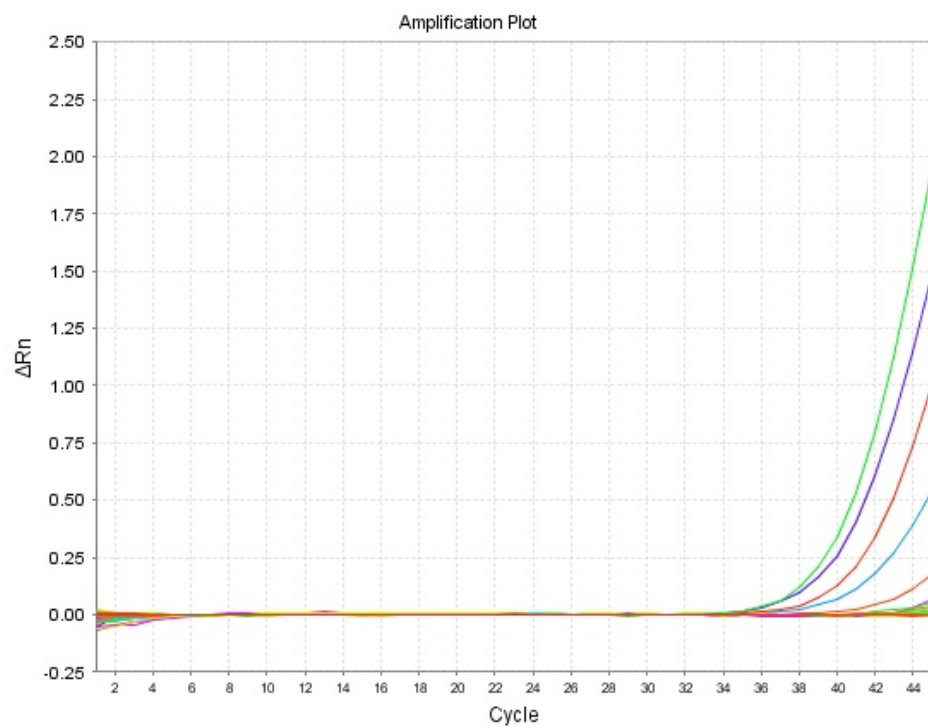


Figure 13 Amplification plot of NOD2bc for phenotypically normal samples.

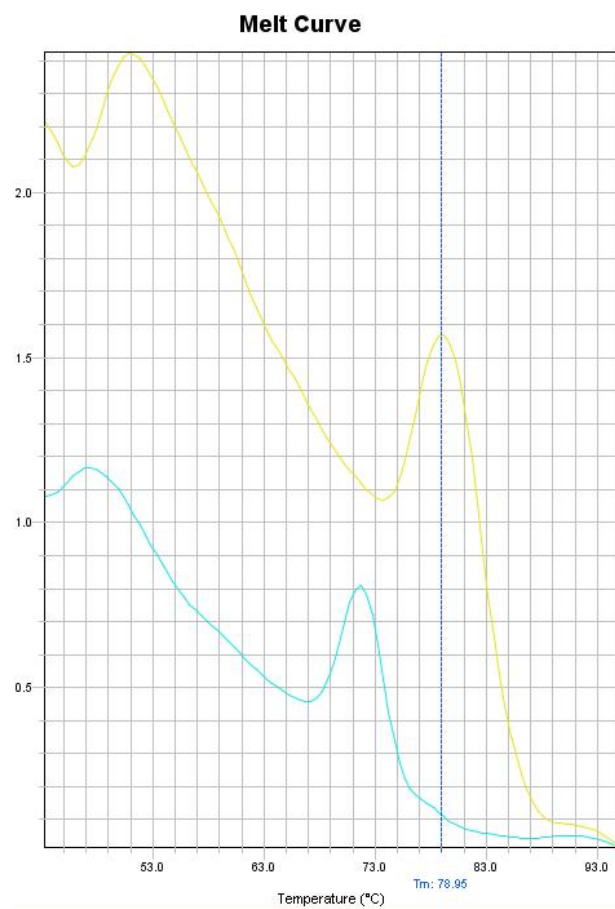


Figure 14 Melt curve of NOD2bc for Crohn's disease samples.

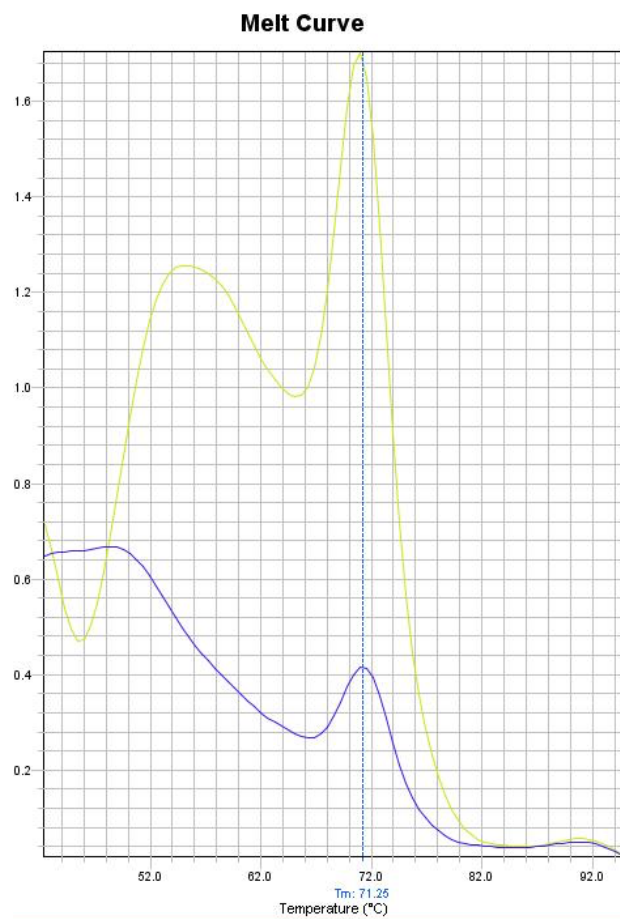


Figure 15 Melt curve of NOD2bc for ulcerative colitis samples.

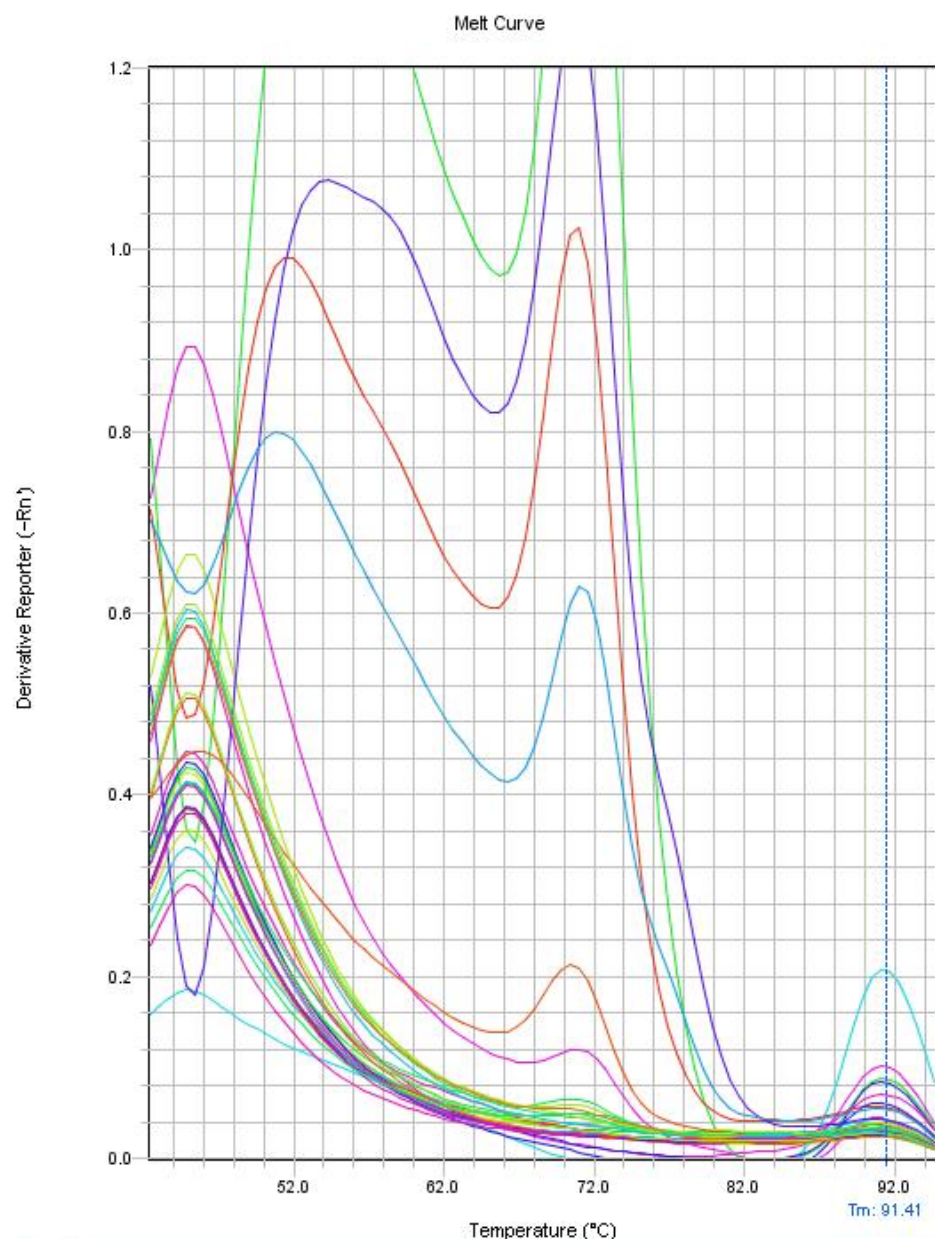


Figure 16 Melt curve of NOD2bc for phenotypically normal samples.



### *Amplification plots and Melt Curves of ATG16L1bc*

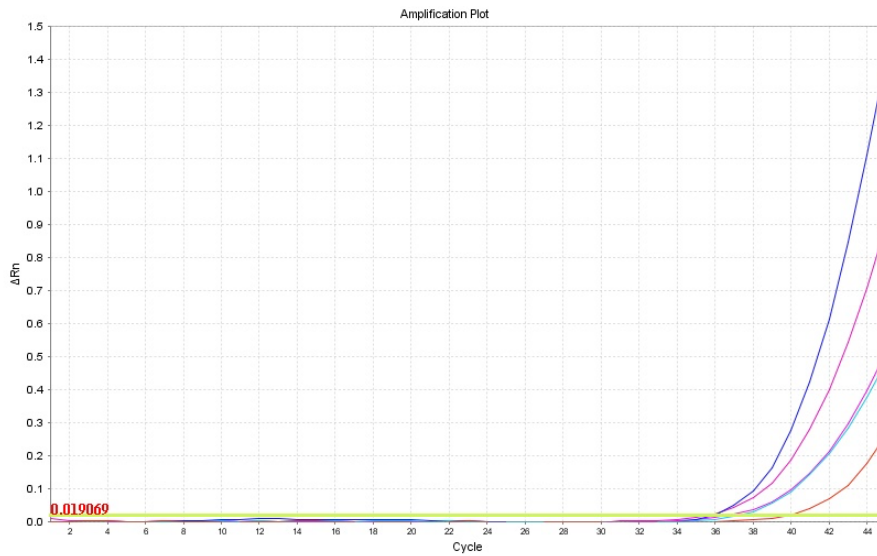


Figure 17 Amplification plot of ATG16L1bc for Crohn's disease samples.

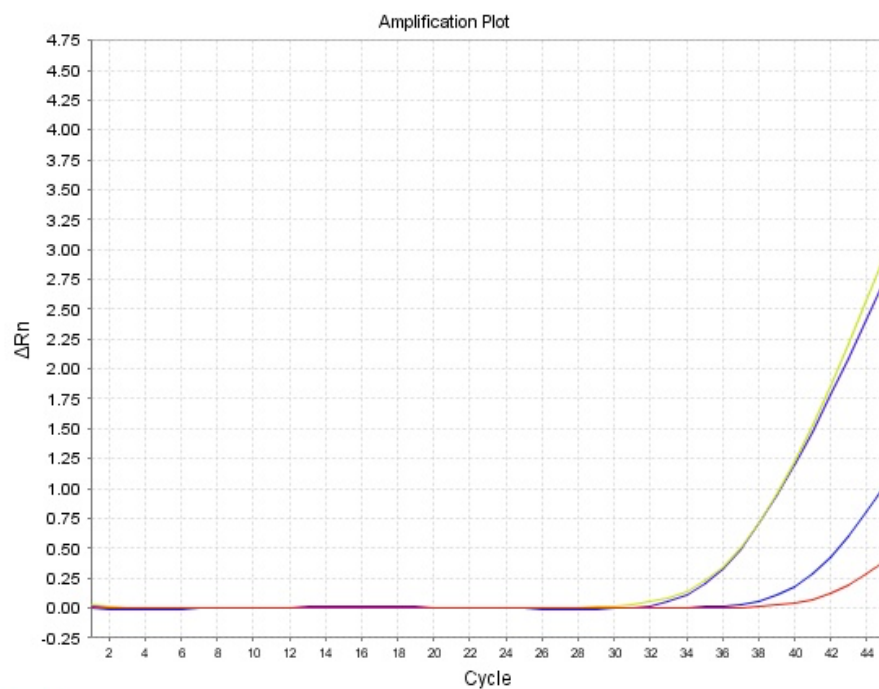


Figure 18 Amplification plot of ATG16L1bc for ulcerative colitis samples.

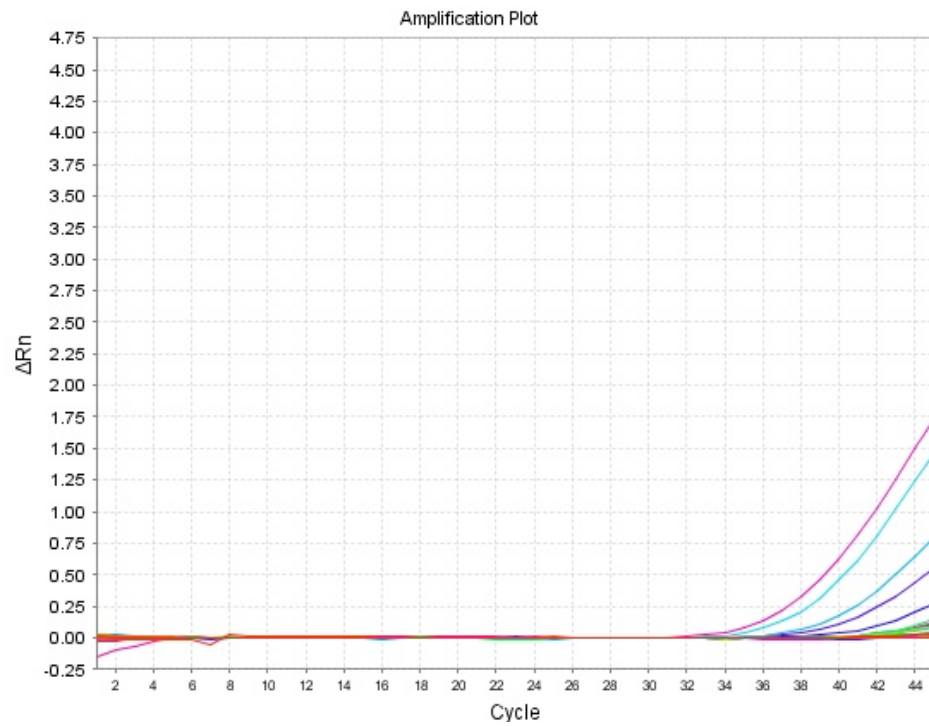


Figure 19 Amplification plot of ATG16L1bc for phenotypically normal samples.

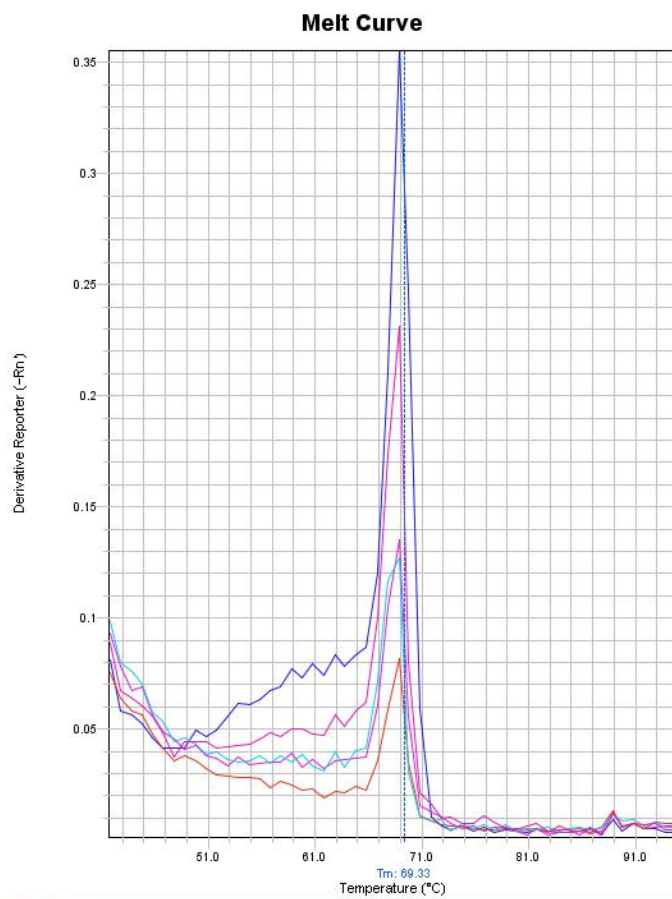


Figure 20 Melt curve of ATG16L1bc for Crohn's disease samples.

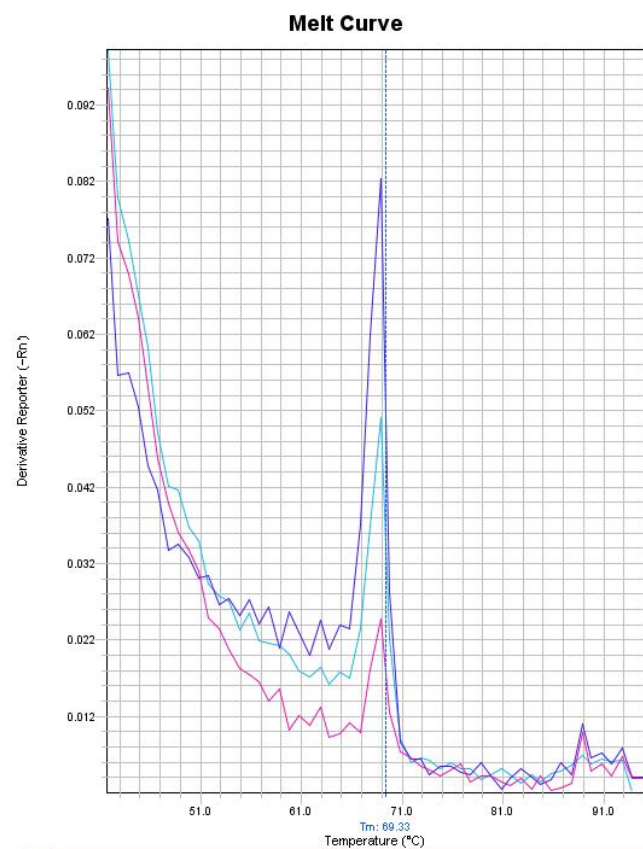


Figure 21 Melt curve of ATG16L1bc for ulcerative colitis samples.

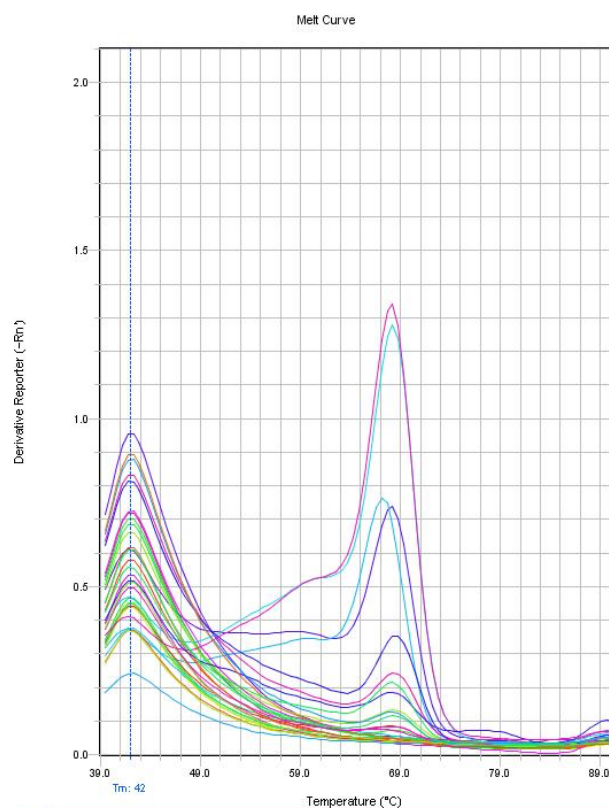


Figure 22 Melt curve of ATG16L1bc for phenotypically normal samples.

## Amplification plots and Melt Curves of PTPN2bc

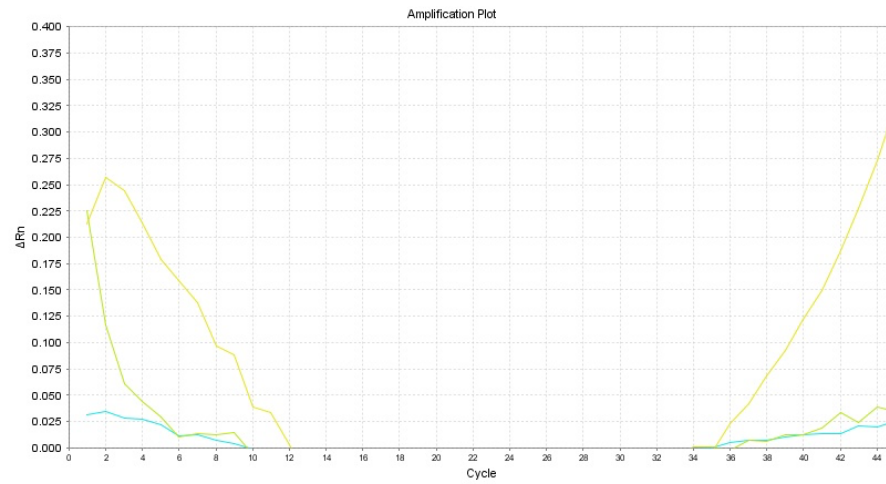


Figure 23 Amplification of PTPN2bc for Crohn's disease samples.

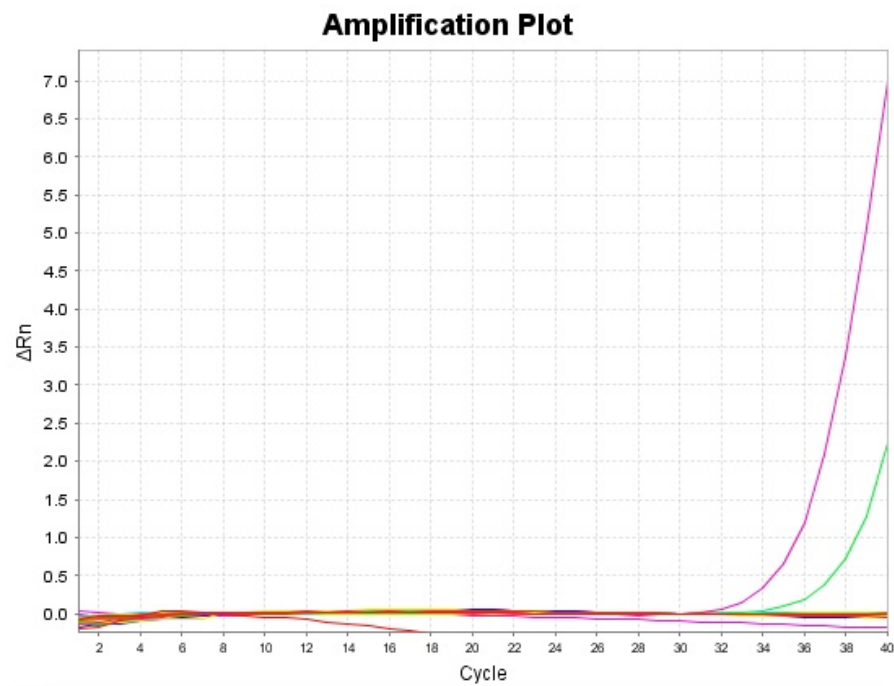


Figure 24 Amplification plot of PTPN2bc for ulcerative colitis samples.

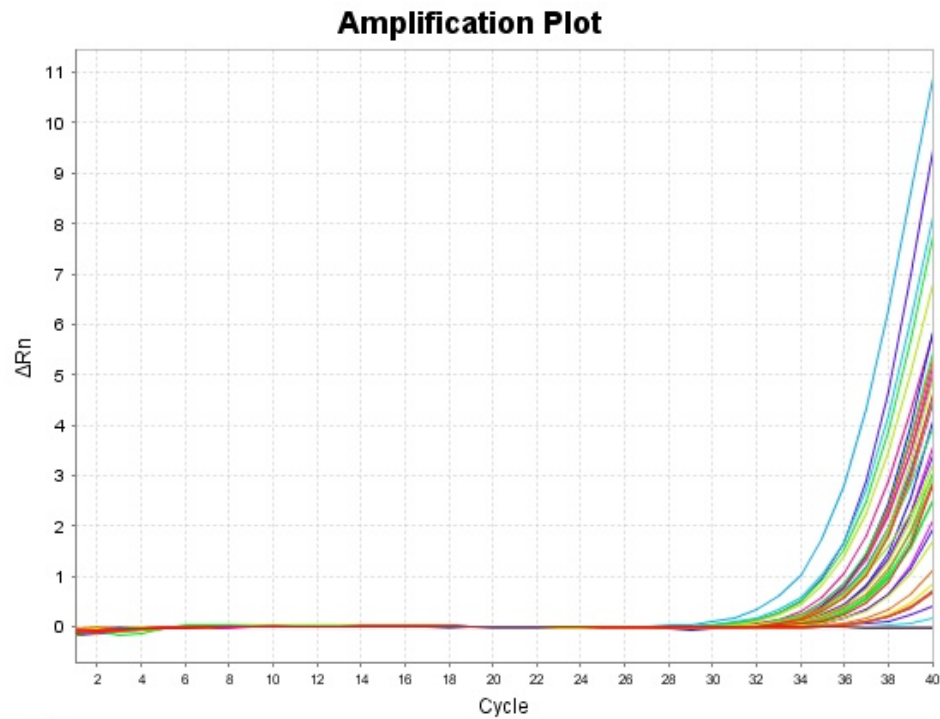


Figure 25 Amplification plot of *PTPN2bc* for phenotypically normal samples.

\*Note: On the PCR run that produced these *PTPN2* amplification plots, I did not ask for a melt curve, only a melt temperature. Thus, I do not have melt curve figures to show for the *PTPN2* gene.

## APPENDIX 5 – INTERVAL PLOTS POST-PCR

### BISULFITE-CONVERTED VARIABLES

Below is a 95% confidence interval plot of each variable.

#### *NOD2bc Interval plots*

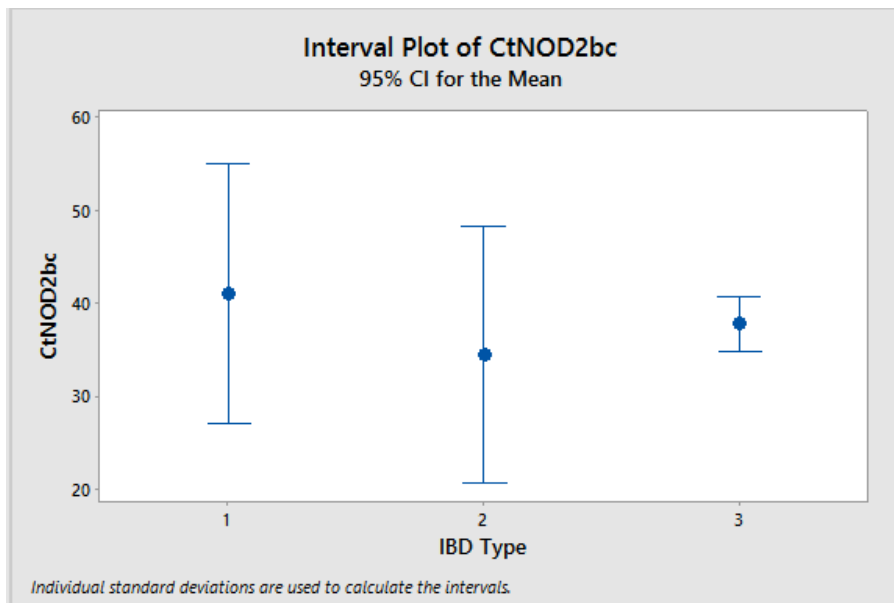


Figure 26 CtNOD2bc Interval plot

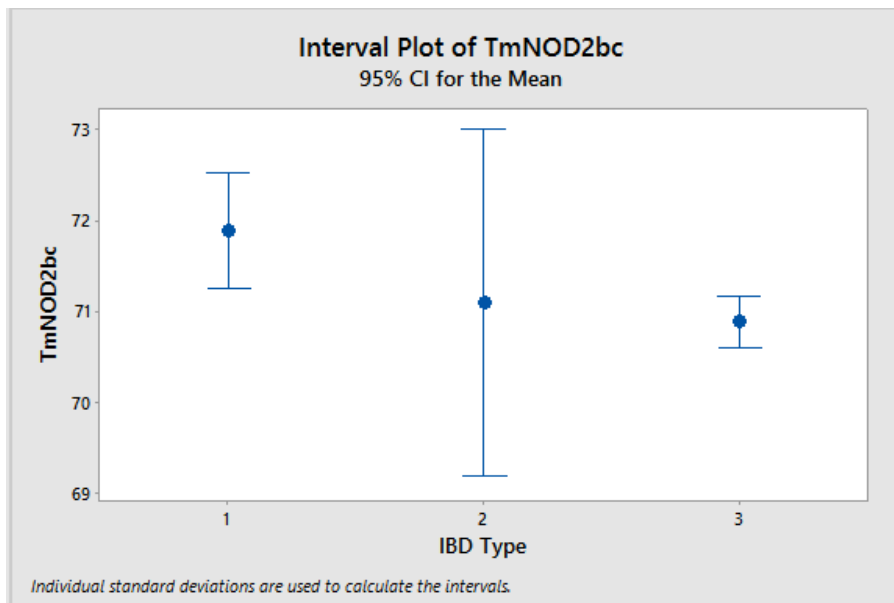


Figure 27 TmNOD2bc Interval plot

#### *ATG16L1bc Interval plots*

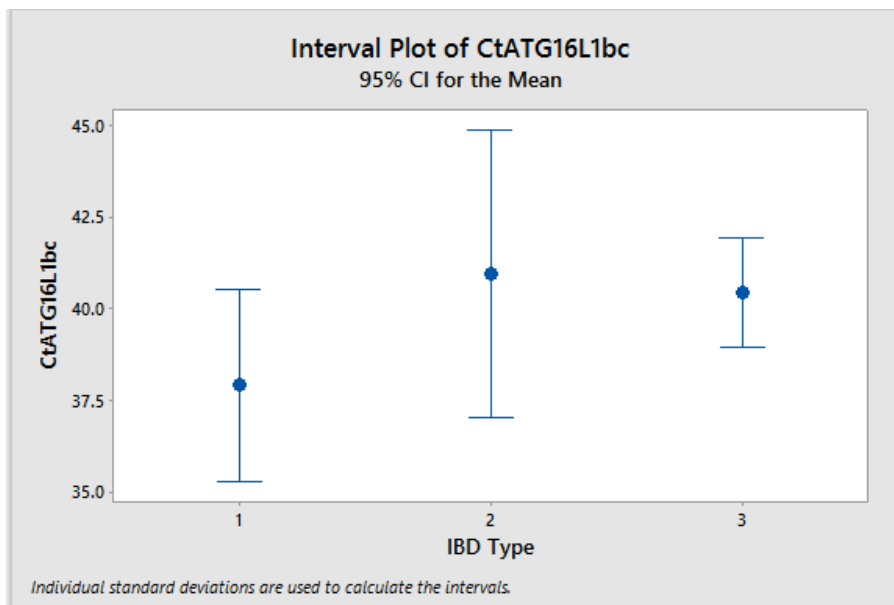


Figure 28 CtATG16L1bc Interval plot

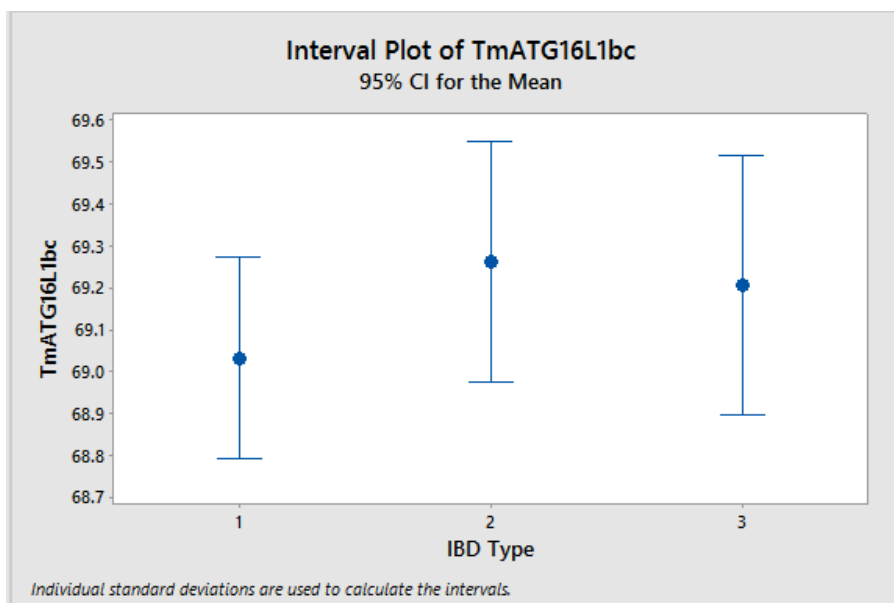
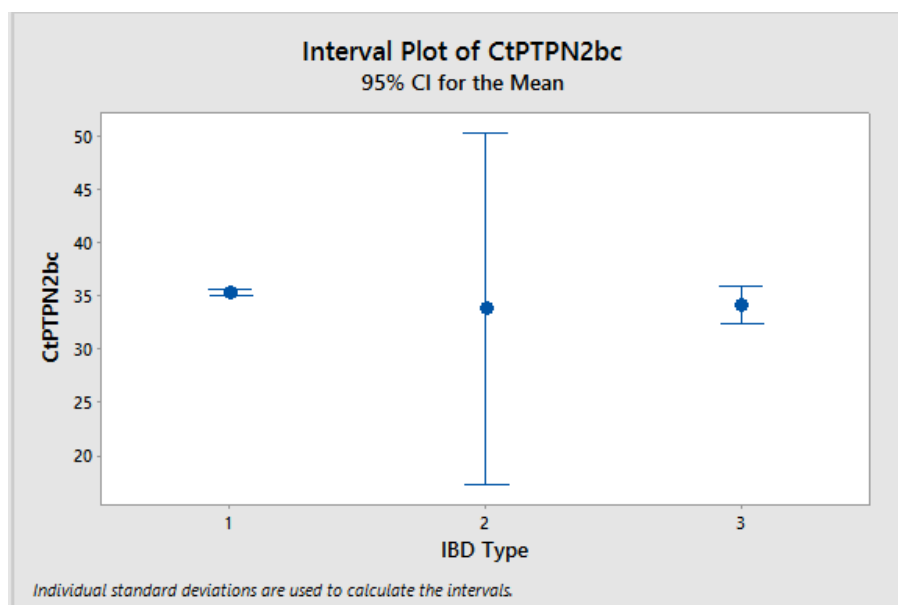
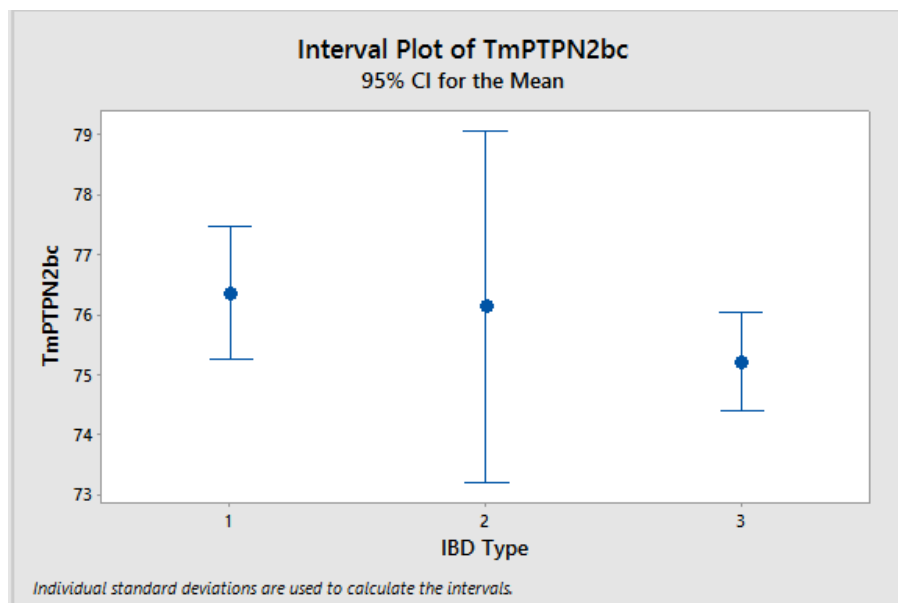


Figure 29 TmATG16L1bc Interval plot

*PTPN2bc Interval plots*



*Figure 30 CtPTPN2bc Interval plot*



*Figure 31 TmPTPN2bc Interval plot*



## APPENDIX 6: CONTINGENCY TABLES WITH ODDS RATIOS, RELATIVE RISKS, AND CONFIDENCE INTERVALS

For each variable, I first weighted the nominal variables by frequencies. I then performed a cross-tabulation, chi-square test and risk estimate. I have separated these tests into comparisons with Crohn's disease, comparisons with ulcerative colitis, and comparisons with IBD. Each table gives the relative risk, the odds ratio, the Fisher's exact test (2-sided) p-value, the chi-square test p-value and the 95% confidence intervals for the factor Crohn's disease (or UC or IBD) / Non-Crohn's disease (or UC or IBD).

*Table 25 Odds ratios, relative risks, and 95% confidence intervals for Crohn's vs non-Crohn's. Note: there were no cases in the yes CD/yes heart attack or celiac disease cell.*

Variable	Relative Risk	Odds Ratio	Fisher's	Chi-square	Lower CI	Upper CI
Allergies	3.613	5.765	0.001	0.000	2.187	15.195
Gender (M/F)	0.951	0.933	1.000	0.890	0.350	2.488
Environment (U/R)	0.767	0.694	0.502	0.427	0.282	1.712
Tobacco Use	1.177	1.253	0.782	0.675	0.436	3.598
Alcohol Use	0.951	0.933	1.000	0.890	0.350	2.488
Familial CD	3.265	8.929	0.010	0.003	1.643	48.510
Heart attack	N/A	N/A	N/A	N/A	N/A	N/A
Cancer	0.303	0.233	0.179	0.140	0.029	1.891
Celiac disease	N/A	N/A	N/A	N/A	N/A	N/A
High cholesterol	0.176	0.125	0.022	0.021	0.016	0.980
Asthma	0.642	0.562	0.726	0.470	0.115	2.737
Rheumatoid arthritis	1.714	2.250	0.375	0.294	0.478	10.595

*Table 26 Odds ratios, relative risks, and 95% confidence intervals for UC vs non-UC. Note: there were no cases in the yes UC/yes familial CD, heart attack, or rheumatoid arthritis cell.*

Variable	Relative Risk	Odds Ratio	Fisher's	Chi-square	Lower CI	Upper CI
Allergies	0.553	0.505	0.389	0.305	0.135	1.896
Gender (M/F)	1.630	1.773	0.339	0.334	0.548	5.730
Environment (U/R)	0.690	0.649	0.575	0.461	0.204	2.061
Tobacco Use	1.700	1.875	0.297	0.318	0.537	6.546
Alcohol Use	0.698	0.661	0.758	0.539	0.175	2.496
Familial CD	N/A	N/A	N/A	N/A	N/A	N/A
Heart attack	N/A	N/A	N/A	N/A	N/A	N/A
Cancer	0.606	0.567	1.000	0.594	0.068	4.701
Celiac disease	1.861	2.148	0.451	0.508	0.212	21.793
High cholesterol	0.353	0.313	0.466	0.250	0.039	2.508
Asthma	0.606	0.567	1.000	0.594	0.068	4.701
Rheumatoid arthritis	N/A	N/A	N/A	N/A	N/A	N/A

Table 27 Odds ratios, relative risks, and 95% confidence intervals for IBD vs non-IBD. Note: there were no cases in the yes UC/yes heart attack cell.

<b>Variable</b>	<b>Relative Risk</b>	<b>Odds Ratio</b>	<b>Fisher's</b>	<b>Chi-square</b>	<b>Lower CI</b>	<b>Upper CI</b>
Allergies	1.828	2.833	0.027	0.015	1.206	6.654
Gender (M/F)	1.150	1.271	0.656	0.590	0.531	3.039
Environment (U/R)	0.740	0.595	0.227	0.210	0.264	1.344
Tobacco Use	1.342	1.684	0.323	0.286	0.642	4.417
Alcohol Use	0.858	0.774	0.658	0.571	0.319	1.880
Familial CD	1.905	4.167	0.111	0.073	0.778	22.321
Heart attack	N/A	N/A	N/A	N/A	N/A	N/A
Cancer	0.404	0.272	0.115	0.085	0.056	1.311
Celiac disease	0.598	0.464	0.642	0.502	0.047	4.580
High cholesterol	0.235	0.133	0.003	0.003	0.029	0.611
Asthma	0.629	0.490	0.356	0.302	0.124	1.940
Rheumatoid arthritis	1.055	1.096	1.000	0.907	0.235	5.103